IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 14 Issue 12

December 2023

ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

SAI

www.ijacsa.thesai.org

Editorial Preface

From the Desk of Managing Editor…

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 14 Issue 12 December 2023
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

CONTENTS

Computer Vision-based Efficient Segmentation Method for Left Ventricular Epicardium and Endocardium using Deep Learning

A F M Saifuddin Saif, Trung Duong, Zachary Holden
Computer Science and Information Systems, West Virginia University Institute of Technology,
Beckley, West Virginia 25801, USA[1]
School of Engineering, Colorado State University Pueblo, Pueblo, CO 81001, USA[2, 3]

***Abstract*—Segmentation of the Left Ventricular Epicardium and Endocardium remains challenging and significant for valuable investigation of cardiac image classification. Previous research methods did not consider the flexibility of the heart area, so measurements needed to be more consistent and accurate. In addition, previous methods ignored the presence of affectability and additional parts, such as the lung organ inside the frame, during segmentation. Deep learning architectures, specifically convolutional neural networks, have become the primary choice for assessing cardiac medical images. In this context, a Convolutional Neural Network (CNN) can be an effective way to segment the left ventricular epicardium and endocardium as CNN can take data pictures, move enormity to various centers or objects in the image and have the choice to separate one from the other. This research proposes an efficient method for segmenting the left ventricular epicardium and endocardium using the InceptionV3 convolutional neural network. Rather than including fully connected layers on the head of the component maps, the proposed method considers the average of each element map, and the subsequent vector was taken care of legitimately into the SoftMax layer. Data augmentation technique was used to validate the proposed method on large number of dataset images. Besides, the proposed method was validated in publicly available MRI cardiac image datasets. Comprehensive experimental analysis was done by analyzing a large number of performance metrics, i.e., cosine similarity, log cos error, mean absolute error, mean absolute percentage error, mean squared error, mean squared logarithmic error, and root mean squared error. The proposed method depicted superior performance for localization of the left ventricular epicardium and endocardium in terms of all these performance metrics. In addition, the proposed method performed efficiently to get smooth curve for covering the region due to usage of interpolation technique to draw the curve, which made it smoother compared with previous research.**

Keywords*—*Convolutional neural network; segmentation; computer vision; deep learning

I. INTRODUCTION

Coronary artery disease (CAD) has the highest morbidity and mortality rates globally [1]. For this, localization of the left ventricular epicardium and endocardium using deep learning approach can be automated to provide a robust tool for imaging the structure of the human heart. Generally, LV segmentation methods are formed based on area and time, where the area locates the heart within the midpoint to the indicated frame [2]. Previous research methods did not consider the flexibility of the heart area, so measurements needed to be more consistent and accurate. Time-based strategies acknowledge the heart to be the central working fact within the frame [3]. These methods endured the absence of affectability and additional parts, such as the lung, which is an active organ inside the frame in expansion for movement production [2]. More estimation has been put forward to handle this issue for LV segmentation in MRI [4] [5]. Besides, due to excellent efficacy or cost ratio in evaluating left ventricular function, gated myocardial perfusion SPECT (MPS) is widely investigated for non-destructive diagnosis of CAD [6]. Endocardium and epicardium must be accurately delineated on perfusion images for quantitative analysis of the left ventricle (LV) in MPS followed by measurement of LV functional parameters. In this context, manual segmentation is time-consuming and needs more reproducibility [6]. Therefore, to improve the accuracy of quantitative analysis, it is necessary to develop a precise, reproducible, and fully automated localization method.

At present, industrial software extracts the left ventricular epicardium and endocardium surface features by estimating the maximum myocardial counts. Then, Gaussian fit is implicated with empirical standard deviation or threshold method to evaluate the details. However, this method needs to be investigated again in assessing myocardial functions. In particular, left ventricular ejection fraction (LVEF) is often overestimated in patients with tiny hearts, and the error is more pronounced in females than males [7].

Traditional computer vision-based image processing methods have demonstrated significant improvement in cardiac image segmentation, such as atlas and model based methods [8, 9]. In recent years, deep learning models, which automatically learn high-level features of the potential distribution of data, outperformed traditional images segmentation methods in accuracy and time efficiency [10]. In this context, multi-class three-dimensional (3D) V-Net was proposed to automatically segment the endocardium and epicardium in gated MPS, which exhibits improved performance [11]. The average Dice similarity coefficient (DSC) values of the model in the endocardium and epicardium of regular patients were 0.907 and 0.965, respectively. However, previous deep learning methods still need to be improved for accurate LV segmentation in MPS towards localization. The object shapes extracted from previous image segmentation methods for localization have shown great success as prior knowledge in refining the deep learning models for medical image classification [12,13]. In this context, combination of prior knowledge reduces the potential output space of model partitioning and speeds up the convergence during the model training. However, prior knowledge is generally used as the input which was hard to extract.

This research uses Inception-v3 CNN architecture for segmenting the left epicardium and endocardium. The innovations and our contributions are listed as follows:

1) This research proposes an efficient method for segmenting the left ventricular epicardium and endocardium using Inception-v3 CNN architecture where the average of each element map and the subsequent vector were taken care of legitimately into the SoftMax layer rather than including fully connected layers on the head of the component maps.

2) Data augmentation technique was used to overcome the shortage of dataset images by the proposed method, which allows the proposed method to be validated robustly.

3) The proposed method was validated by analyzing a large number of performance metrics, i.e., cosine similarity, log cosh error, mean absolute error, mean absolute percentage error, mean squared error, mean squared logarithmic error, and root mean squared error, where proposed method depicted superior performance.

The remainder of this paper is organized as follows: Section I gives the introduction. Critical previous research is illustrated Section II, comprehensive details of the proposed methodology are elaborated in Section III, details of experimental results with analysis for validation are presented in Section IV, and finally, Section V concludes the paper.

II. Previous Research Methods

Manual left ventricular epicardium and endocardium segmentation are crucial for risk stratification, diagnosis, and treatment evaluation. However, manual segmentation has been suffering from various issues, i.e., time-consuming, tedious, and lack of generalization, which can impact the reproducibility of the results [14]. Automatic segmentation for segmentation can overcome some of these limitations [15], where deep learning methods are under investigation to develop accurate, robust, and fast computer vision techniques. As the clinical application of MRI is rapidly growing, robust computer vision techniques are required, which will not need any supervision for acceptable accuracy. This research developed and implemented a robust method for segmenting the Left Ventricular (LV) Epicardium and Endocardium using efficient deep learning method.

Previous Research reported the application of Deep Learning (DL) to segment the left ventricular epicardium and endocardium. Research in [16] evaluated multiple DL methods for left ventricular endocardium segmentation and found the superiority of encoder–decoder-based architectures over non deep learning methods. Research in [17] implicated U-Net to segment the left ventricle by changing UNet architecture in MFP-U-Net. Their proposed CNN added additional convolution layers for producing fixed size feature maps and efficient left ventricular segmentation performance. Research in [18] combined a modified U-Net architecture with an FCN encoder to influence feature extraction and allow the system to learn from execution. Research in [19] implemented bilateral segmentation network to extract deep features and a pyramid local attention algorithm to extract significant features within compact and sparse neighboring contexts. Research in [20] used multiple parallel pipelines for ES and ED frame segmentation using DeepResU-Net. Distinct from the other Research, Research in [22] used self-supervised algorithms [21] to separate the left ventricle to reduce the issue for the lack of labeled data. Research in [23] addressed object detection method and YOLOv3 algorithm, to detect three points of the ventricular chamber and segment the ventricles. Despite these method’s innovativeness and high performance, previous methods focused on segmenting the ventricle for segmentation but not all its anatomical structures.

Convolutional Neural Network (CNN) is a deep learning strategy that can extract data features, move enormity to various centers or objects in the image, and have the choice to separate one from the other. While in harsh strategies for CNN, channels are hand-worked with enough preparation, ConvNets can get capacity with these channels. Most LV limitation techniques are primarily founded on spot-based, time-based, and shape-based speculations, which refer to the areas of strategies except the heart in the picture [24]. A combination of the dynamic figure model and dynamic appearance models were used to confine the left and right ventricles of customary and Tetralogy of Fallot (TOF) hearts on 4-D (3-D+time) MR pictures [25] [26]. For each ventricle, a 4-D model was first used to accomplish incredible essential confinement on all heart stages, and a 3-D model was applied to each stage to increase the exactness while keeping up the complete heartiness of the 4-D division [27]. Another procedure was introduced in Deep CNN to restrict the Left Ventricular in cardiovascular MRI. A six-layered Convolutional Neural Network with different part estimations was used to separate highlights trailed by SoftMax, a connected layer for portrayal.

The pyramids of scales assessment were familiar with the record of the different dimensions of the heart [28]. A range-based device was produced to draw closer to experiencing the ill effects of affectability [29]. Automatic-Image-Driven technique's suppositions depend on the heart, roughly in the middle of the genuine picture. In this context, the LV blood pool is more roundabout than the Right Ventricular blood pool, which has an upper sign force [30]. Artificial Intelligence (AI), Computer Vision (CV), and Image Processing (IP) calculations have been likewise introduced to handle the segmentation of issues by isolating the frontal regional object from the foundation. While hardly any specific methods have been proposed to deal with the issue of LV restriction in X-beam, a couple of computer vision and image processing methods have been familiar with limited unmistakable body parts in modalities, i.e., Ultrasound and Computed Tomography (CT). Kellman assessed and limited the LV posture using probabilistic boosting trees and minimal space learning [31]. Research in [32] utilized nonlinear planning through relapse to limit in echodiogramic dataset. Modified LV (Left Ventricle) limitation in cardiovascular MRI pictures is a significant development for programmed division and practical perception. In this context, a comparative investigation should be combined into the severe degree of chance for backslide to improve the restriction task [33]. Recently, substantial convolutional systems have accomplished magnificent execution in many pictorial division assignments [34] and are excitedly applied in the field of clinical picture appraisal [35]. For instance, research in [36] suggested a 3D essentially oversaw system for the robotized division of the liver and the entire heart, which needs further investigation due to a lack of datasets. Different study attempts were made to approach the problem of segmentation of left ventricular epicardium and endocardium activity. This research proposes an efficient method for segmenting the left ventricular epicardium and endocardium using the InceptionV3 CNN model. The proposed method considers the average of each element map, and the subsequent vector was taken care of legitimately into the SoftMax layer. The proposed method depicted superior performance for segmentation of left ventricular epicardium and endocardium.

III. Proposed Research Methodology

This research used InceptionV3 as the CNN architecture. The overall proposed methodology is shown in Fig. 1.

Top layer with a custom network is trained rather than including fully connected layers on the head of the component maps. The average of each element map was used, and the subsequent vector was taken care of legitimately into the SoftMax layer. Global average pooling was aggregated with spatial data for spatial solid interpretations of the features. The input layer was normalized by adjusting and scaling the activations. To build the solidness of the network, batch normalization normalized the yield of a previous actuation layer by removing the batch mean and separating it by batch standard deviation.

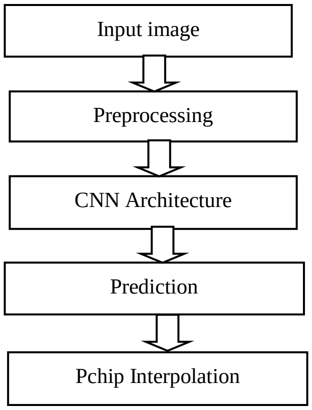


Fig. 1. Proposed methodology.

A. Input Images

Cardiac MR Image sequences with short-axis were used from 33 subjects for 7980 2D images. All the subjects were under the age of 18 where each patient's image sequence consisted of exactly 20 frames, and the number of slices collected along the long axis of the subjects ranges between 8 and 15. Spacing between slices ranged between 6 and 13 mm. Each image slice consisted of 256 * 256 pixels with a pixel spacing of 0.93–1.64 mm.

B. Preprocessing

Each subject's arrangement was comprised of 20 frames and 8 to 15 slices along the long axis for an aggregate of 7980 pictures. However, images were raw and unprocessed, kept as 16-bit DICOM images. So, this research converted 16-bit DICOM input images into 256*256*3 by reshaping them, as shown in Fig. 2.

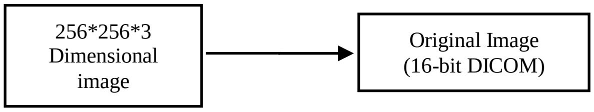


Fig. 2. Conversion of 16-bit DICOM input images into 256*256*3.

C. Data Augmentation

The paucity of data is another main problem in establishing deep learning models like CNN. Data augmentation is a helpful strategy in building a convolutional neural network that can expand the size of the training set without procuring new pictures. In this context, frames are copied with some variety. This research expanded the image to safeguard the highlights key to make predictions yet revamp the pixels enough that it includes some noise. In addition, this research rescaled images by dividing 255 with every pixel. In this context, insufficient data for model training was a typical scenario, whereas 5011 segmentation images were available. The data augmentation technique provided strong support with reduced loss in that context.

D. CNN Architecture

Each input image was passed through convolution layers with kernels, max pooling, and fully connected layers. SoftMax function was applied to segment objects with probabilistic values between 0 and 1. Convolution preserved the correlation between pixels by understanding features using squares of input data. Max pooling is considered the most significant element from the dense feature map. Fully connected layers are used where all the inputs from one layer are added to every activation unit of the next layer. In this context, the last few layers were fully connected layers, which compiled the data extracted by previous layers to form the final output. CNN architecture deployed by this research is shown in Fig. 3.

E. Prediction

The cardiac MRI dataset provides short-axis cardiac MR images and ground truth of their left ventricles endocardial and epicardial segmentations. Each image was manually segmented for a total of 7980 images where both the endocardium and epicardium of the left ventricle were visible, for a total of 5011 segmented MR images and 10022 contours. Each contour was described by 32 points given in pixel coordinates. This research trained the proposed method based on these 32 points (target variable). After training the model, a prediction of 64 points (32 for epicardium and 32 for endocardium) was acquired.

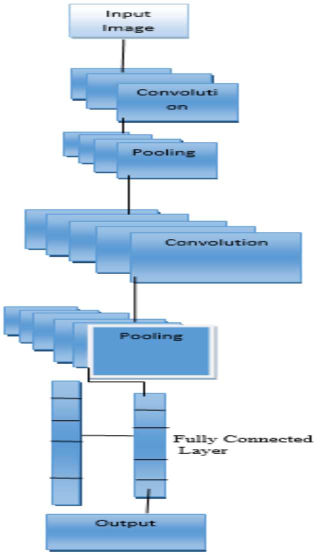


Fig. 3. CNN architecture used in the proposed research.

F. Interpolation

The proposed method used interpolation to draw the smooth curve. In this context, for each image, the epicardium and endocardium contour were depicted by 32 points in pixel coordinates (so, in total, 64 points). In this context, Pchip stands for Piecewise Cubic Hermite Interpolating polynomial used by the proposed method that interpolates data and specified derivatives at the interpolation points. As two points determine a linear function, two points and two given slopes determine a cubic. The data points are known as "knots." Y-values remain at the knots, so to get a particular PCHIP, the proposed method specified the values of the derivative y at the knots. In addition, these two cubic polynomials were considered in x on the interval 1⩽x⩽2. These functions were formed by adding cubic terms that vanish at the endpoints to the linear interpolant. After getting the predicted values of epicardium and endocardium, those points were passed for interpolation.

IV. Experimental Results and Discussion

A. Dataset

Cardiac MRI datasets [30] were developed from 33 subjects. This dataset contains 20 frames and 8-15 slices along the axis for individual subject sequences. Most slices in this dataset contain heart anomalies because of some heart diseases. So, this dataset consists of a total of 7980 images. In Cardiac MRI datasets, every patient contains 20 timeframes for 8-15 short-axis slices with matrix size 256x256, 6-13 mm slice thickness, and pixel resolution of 0.93-1.64mm [31]. The images were originally stored as 16-bit DICOM images. Form 7980 images, there were 5011 segmented images with 10022 contours. Each shape was pointed by 32 given in pixel coordinates.

In Cardiac MRI datasets, 4008 images were used by this research for training purposes. This research used 501 images for validation, and for testing, this research used 502 images. Besides, "Adam" as an optimization algorithm was used by the proposed method [37][38]. The preliminary learning rate was 0.02, the learning rate decreased with factor = 0.4, and patience was three by monitoring the 'validation loss.' Relu was used as an activation function and in the outcome layer, proposed used liner as an activation function.

B. Hardware and Software Set Up

For experimental purposes, a Windows platform was used with an 8th generation Intel Quad-Core i7-7300HQ processor (14MB Cache, 4.0GHz), 16GB DDR4 DRAM, and NVIDIA GeForce GTX 1050 with 16GB VRAM. This research used TensorFlow v2.13.0 and Keras 2.13.1 RC1 [39] [40]. This research also used some modules, i.e., Numpy, Matplotlib, Itk, Seaborn, Sklearn, and Scipy, for experimentation. The whole dataset was divided into two sections, i.e., one for training and the other for testing. The training section contains 80 % of the dataset images, and the remaining 20% was used for testing.

C. Performance Metrics

To validate the performance of the proposed method, various performance metrics were used, i.e., cosine similarity, log cosh error, mean absolute error, mean absolute percentage error, mean squared error, mean squared logarithmic error, and root mean squared error for graphical representation. Details of these metrics are mentioned in subsequent sections.

1) Cosine similarity: Estimating the comparability between at least two vectors is called cosine similarity [41]. The vectors are ordinarily nonzero and are inside an inward item space. In practice, cosine similarity is used to decide how comparable the documents are regardless of their size, which is a value equal to the division between the dot product of vectors and the product of the Euclidean norms or magnitude of each vector mentioned in Eq. (1).

$$similarity = \cos\theta = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \quad (1)$$

Here, A and B are two vectors. A.B is the dot product of those two vectors.

2) Log cosh error: Log Cosh Error is used in a regression task, a logarithm of hyperbolic cosine of the prediction error. Log Cosh Error works like mean squared error, which is not firmly influenced by the occasional wildly incorrect prediction. Log Cost Error was estimated using Eq. (2).

$$L(y, y^p) = \sum_{i=1}^{n} \log(\cosh(y_i^p - y_i)) \quad (2)$$

In this equation, L denotes Log Cosh Error, log () denotes logarithm function, cosh() denotes hyperbolic cosine function, $(y_i^p - y_i)$denotes difference between two points in y-axis.

3) Mean Absolute Error (MAE): Mean Absolute Error (MAE) is a widely recognized metric that quantifies precision for persistent factors [42]. MAE estimates the average magnitude of the errors in a set of predictions without thinking about their path. Besides, MAE is the difference between true or actual values and predicted values measured using Eq. (3).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_j - \hat{y_j}\right| \tag{3}$$

Here, MAE denotes Mean Absolute Error, $y_j - \hat{y_j}$ denotes difference between prediction and true value, n denotes total number of data points.

4) Mean absolute percentage error: Mean Absolute Percentage Error (MAPE) is the mean or normal of the total rate errors of forecasts [43]. Error is characterized as the result of the observed value by subtracting the forecasted value. Percentage errors are added regardless of the sign to register MAPE. This measurement is straightforward because it gives the error as far as percentage. Likewise, absolute percentage errors are utilized, and the issue of positive and negative errors offsetting each other is dodged. Thus, MAPE has administrative allure and is regularly utilized in estimation. MAPE can be well defined by middling the Absolute Percentage Errors of forecasts. MAPE is estimated using Eq. (4).

$$MAPE = \{(Estimated\ Value - Actual\ Value) \times 100\} \tag{4}$$

5) Mean Squared Error (MSE): The average of the squared error is called the Mean Squared Error (MSE) [44]. MSE reveals how close a regression line is to a set of points, which is done by taking the distances from the points to the regression line. The squaring is essential to eliminate any negative signs, which likewise gives more weight to bigger contrasts using Eq. (5).

$$MSE = \frac{1}{y}\sum_{i=1}^{n}\left|y_j - \tilde{y_i}\right| \tag{5}$$

Here, MSE denotes Mean squared error, n denotes number of data points, y_j denotes observed values and $\tilde{y_i}$ denotes predicted values.

6) Root Mean Squared Logarithmic Error (RMSLE): Root-mean-squared-logarithmic error (RMSLE) is a function mentioned in Eq. (6) used for finding the difference between predicted values and the actual values [45]. To comprehend the manipulation of RMSLE and corresponding disparities, it is imperative to calculate means squared Error (MSE) mean where MSE joins both fluctuation and inclination of the indicator. MSLE cares about the relative difference between predicted value and actual value.

$$RSMLE = \{\log(prediction + 1) - \log(actual + 1)\}^2 \tag{6}$$

7) Root Mean Squared Error (RMSE): Root Mean Squared Error (RMSE) is a recognized way to find the error of a regression model, which defines how close a fitted line is to data points [44]. RMSE was estimated using Eq. (7).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x_i})}{N}} \tag{7}$$

Here, i denotes variable, N denotes number of non-missing data points, x_i denotes actual observed values and $\hat{x_i}$ denotes estimated time series.

D. Experimental Results

The proposed method received the best cosine similarity of 0.9977 for the Cardiac MRI dataset after 76 epochs, as shown in Fig. 4. For each of the performance metrics, the graphical representation is shown in Fig. 4 to 9, where the orange line was used for training, and the blue line was used for validation. This research considered epoch along with the x-axis and cosine similarity along with the y-axis. Estimation of cosine similarity turned into an essential factor for understanding likenesses between objects and provided the strong assumption between the train set and validation set's similarity. In addition, loss differences can be calculated from this estimation, which helped to ensure that the proposed method never faces an overfitting problem.

Cosine similarity calculates the comparability between two vectors of an inner product space. The output produces a value ranging from -1 to 1, indicating similarity where -1 is non-similar, 0 is orthogonal (perpendicular), and 1 represents total similarity. From Fig. 4, it can be observed that in 76 epochs, the value of cosine similarity was 0.9977, and then the model stopped learning. If the model continued learning, the proposed method would face the overfitting problem.

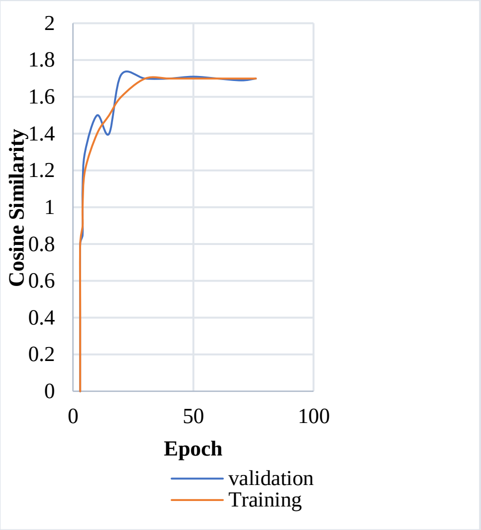


Fig. 4. Cosine similarity of cardiac MRI dataset.

This research received Log Cosh Error of 1.4585. Log Cosh Error was strongly affected by the occasional wide incorrection prediction and thus considered an improved version of MSE. Log Cosh Error is preferable to use when dataset contains significant errors due to more sensitivity to errors than the MSE. For this reason, this research used Log cos error for validating the proposed method. From Fig. 5, in 76 epochs, the model stopped learning. From 1 to 20 epoch, it had a large loss difference between train and validation data. But after the 20th epoch, loss difference was less.

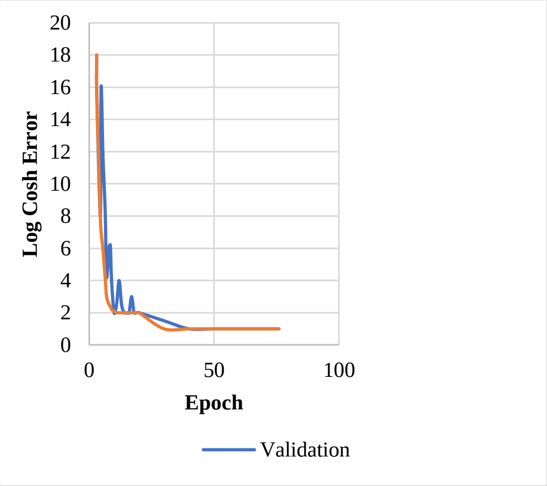


Fig. 5. Log cosh error of cardiac MRI dataset.

This research received Mean Absolute Error (MAE) of 2.0178. In many regression scenarios, Mean Absolute Error is preferable when the average error becomes very large, which is the main reason to use MAE in this research for validating the proposed method. In Fig. 6, in 76 epochs, MAE was 2.0178, then the model stopped learning. From 1 to 25 epochs, a significant loss difference was observed between train and validation data. However, after the 25th epoch, the loss difference was less.

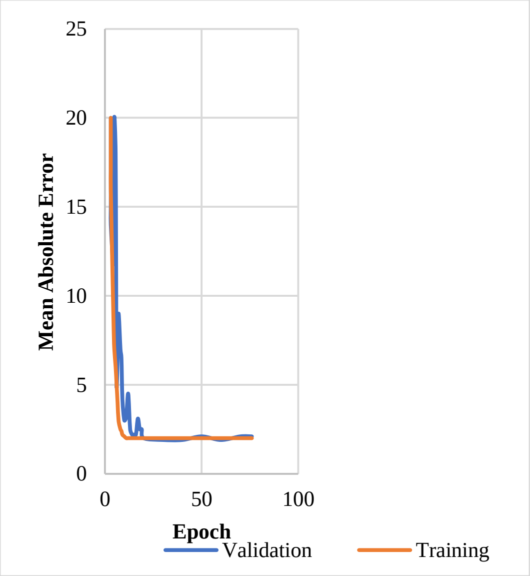


Fig. 6. Mean absolute error of cardiac MRI dataset.

This research received Mean-Squared-Logarithmic Error (MSLE) of 4.1755e-04. Usage of MSLE during regression prevented significant errors from being significantly more penalized than small ones. For cases where the target value range was large, this was the main reason to use MSLE for validating the proposed method. In Fig. 7, in 76th epochs learning was stopped where from 1 to 20th epoch, the proposed method had a significant loss difference between train and validation data. However, after 20 epochs, the loss difference was less.

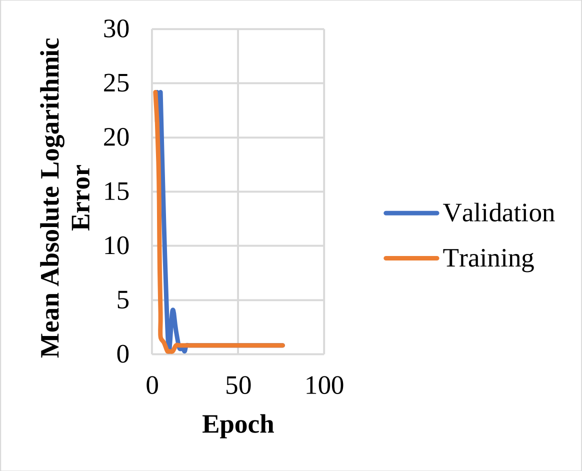


Fig. 7. Mean absolute percentage error on cardiac MRI dataset.

This research received Mean Squared Error (MSE) of 7.0577. MSE is preferable to use when the average error is very small. One minor difference with MAE was that the result is squared, which introduced benefits during optimization, which was the main reason for this research to use MSE for validation. In Fig. 8, in 76 epochs, the result of the Mean Squared Error was 7.0577, then the model stopped learning. From 1 to 25 epochs, it had a significant loss difference between train and validation data. However, after 25 epochs, the loss difference was less.

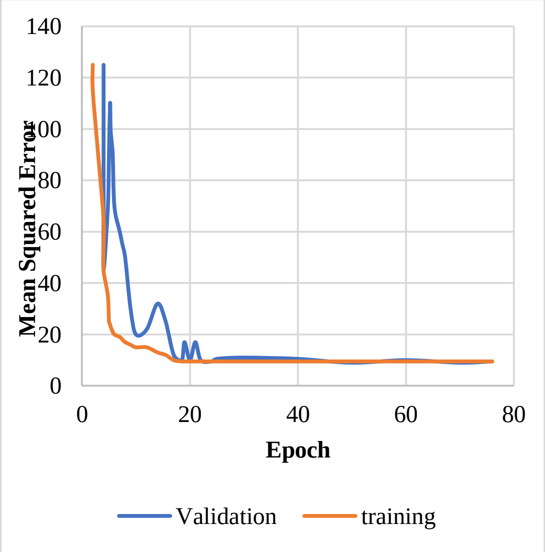


Fig. 8. Mean squared error of cardiac MRI dataset.

The proposed method received Root Mean Squared Error (RMSE) of 2.6561. RMSE was mostly applicable when significant errors were undesirable, which is the main reason for using RMSE in this research for validation. In Fig. 9, in 76 epochs, the Root Mean Squared Error result was 2.6561, then the model stopped learning. From 1 to 20 epochs, it had a significant loss difference between train and validation data. However, after 20 epochs, the loss difference was less. Overall experimental results for the proposed method are in Table I. Sample resultant outputs as image our show in Fig. 10.

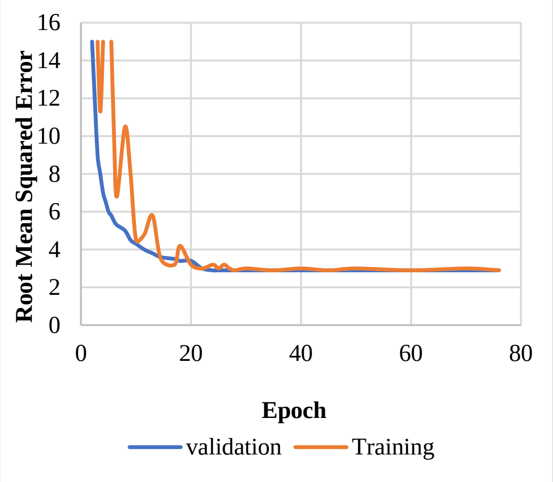
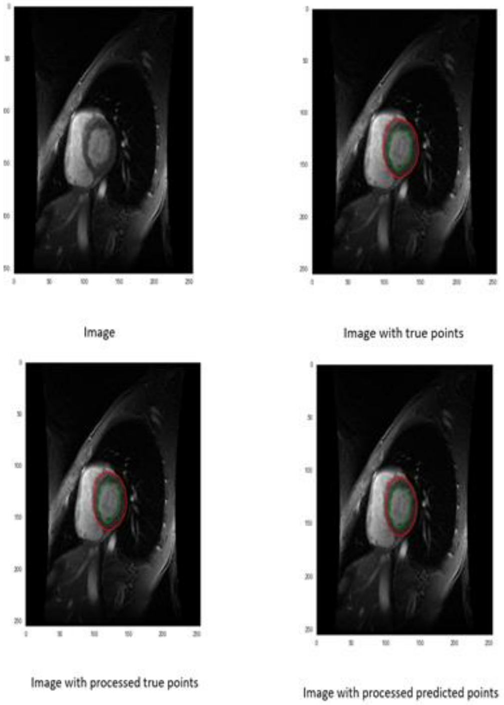


Fig. 9. Root mean square error of cardiac MRI dataset.

TABLE I. OVERALL EXPERIMENTAL RESULTS BASED ON CARDIAC MRI DATASET

Performance Metrics	Result
Cosine similarity	0.9977
Log Cosh Error	1.4585
Mean Absolute Error	2.0178
Mean Squared Logarithmic Error	4.1755e-04
Mean Squared Error	7.0577
Root Mean Squared Error	2.6561
Training Processing Time in Seconds	2500 s



Image

Image with true points

Image with processed true points

Image with processed predicted points

Fig. 10. Same output from Cardiac MRI datset.

The proposed method was compared with existing research based on several metrics, i.e., Mean Absolute Error (MAE), Mean Absolute Percentage Error and Root Mean Squared Error (RMSE). Mean Absolute Error (MAE) was used for summarizing and measuring the quality of a deep learning model. The proposed method received MAE 2.0178, shown in Table II, after the compilation of training. Previously, research in [46] estimated the mean absolute error of 2.34 using a lightweight left ventricle localizer approach. The performance difference between the research in [46] and the proposed method by this research is that the number of data presented in the dataset used by research in [46] was significantly low to train the model perfectly. This research overcame the obstacle by using data augmentation approach to balance the sample data. Data augmentation is a significant strategy in building a convolutional neural network that can expand the size of the training set without procuring new frames. The proposed method rescaled the image by dividing 255 with every pixel. The proposed research could feed a decent amount of data through the network through this approach.

TABLE II. COMPARISON BASED ON MEAN ABSOLUTE ERROR

Method	Mean Absolute Error
Proposed Inception-V3 Convolutional Neural Network	2.0178
LVLNET and Fully Convolutional Neural Network [46]	2.34

The proposed method received Mean Absolute Percentage Error value of 1.5661, shown in Table III. Research in [47] received Mean Absolute Percentage Error of 1.43 using 3D active appearance models (AAM) [47]. Compared with the proposed method by this research with the 3D active appearance models on short axis cardiac, the proposed method provided better performance in terms of Mean Absolute Percentage Error. Research in [47] used the Gauss-Newton optimization technique. In contrast, this research did not use the Gauss-Newton optimization technique because even though Gauss-Newton optimization is accurate and reliable, the Gauss-Newton optimization technique is slow.

TABLE III. COMPARISON BASED ON MEAN ABSOLUTE PERCENTAGE ERROR

	Mean Absolute Percentage Error
Inception-V3 Convolutional Neural Network	1.5661
3D Active Appearance Model (AAM) and 2D + time active shape model (ASM) [47]	1.43

Root Mean Squared Error (RMSE) is primarily useful when significant errors are particularly undesirable. For RMSE, the lower value is better. In the proposed method, this research received RMSE of 2.6561. By using the multiple linear regression (MLR) model, research in [48] received an RMSE of 6.24. Using the random-forest regression (RFR) model, research in [46] received RMSE of 5.72. So, in both cases, this research's proposed method received better results compared with research in [46] and [48].

V. Conclusion

This research proposed an efficient method to segment left ventricular epicardium and endocardium using convolutional neural network. Data augmentation technique implicated by the proposed method expanded the size of the training set without procuring new images thus assisted to validate the proposed methos on large number of dataset images. This research used Cardiac MRI dataset to validate the proposed methodology and estimated various performance metrics to justify the effectiveness robustly. After reshaping and rescaling the datasets images, proposed method used Inception V3 without top layer. Pchip interpolation technique was applied to smoothen the curve in the output images. This research observed the loss 2.011 because of the paucity of data which indicates effective localization of epicardium and endocardium with less error. Besides, due to usage of data augmentation, proposed research was able to feed a decent amount of data through network comparing with existing research. In addition, in comparison with other research, for 3D active appearance models on short axis cardiac, the proposed method provided better performance. In future, this research aims to improve CNN framework using various recent deep learning architecture such as Faster R-CNN and YOLOv7 for vast varieties of dataset images using data augmentation technique. Proposed method is expected to contribute significantly for more precise discovery and treatment of cardiovascular disease.

Acknowledgment

Authors would like to thank West Virginia University Institute of Technology and School of Engineering, Colorado State University Pueblo for various support for this research.

References

[1] A. Garavand, A. Behmanesh, N. Aslani, H. Sadeghsalehi, and M. Ghaderzadeh, "Towards diagnostic aided systems in coronary artery disease detection: a comprehensive multiview survey of the state of the art," *International Journal of Intelligent Systems,* vol. 2023, 2023.

[2] M. A. Shoaib, J. H. Chuah, R. Ali, S. Dhanalakshmi, Y. C. Hum, A. Khalil, and K. W. Lai, "Fully Automatic Left Ventricle Segmentation Using Bilateral Lightweight Deep Neural Network," *Life,* vol. 13, p. 124, 2023.

[3] W. H. Marshall, "Development of a competency-based curriculum in advanced heart failure training for the adult congenital heart disease cardiology fellow," The Ohio State University, 2023.

[4] N. Das and S. Das, "A review on right ventricle cardiac MRI segmentation," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* vol. 11, pp. 1348-1358, 2023.

[5] S. H. Chuah, L. K. Tan, N. A. Md Sari, B. T. Chan, K. Hasikin, E. Lim, N. M. Ung, Y. F. Abdul Aziz, J. Jayabalan, and Y. M. Liew, "Remodeling in Aortic Stenosis With Reduced and Preserved Ejection Fraction: Insight on Motion Abnormality Via 3D+ Time Personalized LV Modeling in Cardiac MRI," *Journal of Magnetic Resonance Imaging,* 2023.

[6] F. Zhu, L. Li, J. Zhao, C. Zhao, S. Tang, J. Nan, Y. Li, Z. Zhao, J. Shi, and Z. Chen, "A new method incorporating deep learning with shape priors for left ventricular segmentation in myocardial perfusion SPECT images," *Computers in biology and medicine,* vol. 160, p. 106954, 2023.

[7] K. Nakajima, T. Shibutani, F. Massanes, T. Shimizu, S. Yoshida, M. Onoguchi, S. Kinuya, and A. H. Vija, "Myocardial perfusion imaging with retrospective gating and integrated correction of attenuation, scatter, respiration, motion, and arrhythmia," *Journal of Nuclear Cardiology,* pp. 1-17, 2023.

[8] X. Shu, Y. Yang, J. Liu, X. Chang, and B. Wu, "ALVLS: Adaptive local variances-Based levelset framework for medical images segmentation," *Pattern Recognition,* vol. 136, p. 109257, 2023.

[9] F. Yuan, Z. Zhang, and Z. Fang, "An effective CNN and Transformer complementary network for medical image segmentation," *Pattern Recognition,* vol. 136, p. 109228, 2023.

[10] Z. R. Mahayuddin and A. Saif, "A comprehensive review towards segmentation and detection of cancer cell and tumor for dynamic 3D reconstruction," *Asia-Pacific Journal of information technology and multimedia,* vol. 9, pp. 28-39, 2020.

[11] Y. Zhang, F. Wang, H. Wu, Y. Yang, W. Xu, S. Wang, W. Chen, and L. Lu, "An automatic segmentation method with self-attention mechanism on left ventricle in gated PET/CT myocardial perfusion imaging," *Computer Methods and Programs in Biomedicine,* vol. 229, p. 107267, 2023.

[12] S. Barraza-Aguirre, J. Diaz-Roman, C. Ochoa-Zezzatti, B. Mederos-Madrazo, J. Cota-Ruiz, and F. Enriquez-Aguilera, "Segmentation of Lung Lesions Caused by COVID-19 in Computed Tomography Images Using Deep Learning," in *Internet of Everything for Smart City and Smart Healthcare Applications,* ed: Springer, 2023, pp. 237-259.

[13] C. Zhao, Y. Xu, Z. He, J. Tang, Y. Zhang, J. Han, Y. Shi, and W. Zhou, "Lung segmentation and automatic detection of COVID-19 using radiomic features from chest CT images," *Pattern Recognition,* vol. 119, p. 108071, 2021/11/01/ 2021.

[14] N. Vladimirov, E. Brui, A. Levchuk, W. Al-Haidri, V. Fokin, A. Efimtcev, and D. Bendahan, "CNN-based fully automatic wrist cartilage volume quantification in MR images: A comparative analysis between different CNN architectures," *Magnetic Resonance in Medicine,* 2023.

[15] H. Dang, M. Li, X. Tao, G. Zhang, and X. Qi, "LVSegNet: A novel deep learning-based framework for left ventricle automatic segmentation using magnetic resonance imaging," *Computer Communications,* vol. 208, pp. 124-135, 2023.

[16] C. D. Reddy, L. Lopez, D. Ouyang, J. Y. Zou, and B. He, "Video-Based Deep Learning for Automated Assessment of Left Ventricular Ejection Fraction in Pediatric Patients," *Journal of the American Society of Echocardiography,* vol. 36, pp. 482-489, 2023.

[17] S. Moradi, M. G. Oghli, A. Alizadehasl, I. Shiri, N. Oveisi, M. Oveisi, M. Maleki, and J. Dhooge, "MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography," *Physica Medica,* vol. 67, pp. 58-69, 2019.

[18] K. B. Girum, G. Créhange, and A. Lalande, "Learning with context feedback loop for robust medical image segmentation," *IEEE transactions on medical imaging,* vol. 40, pp. 1542-1554, 2021.

[19] F. Liu, K. Wang, D. Liu, X. Yang, and J. Tian, "Deep pyramid local attention neural network for cardiac structure segmentation in two-dimensional echocardiography," *Medical image analysis,* vol. 67, p. 101873, 2021.

[20] M. G. R. Alam, A. M. Khan, M. F. Shejuty, S. I. Zubayear, M. N. Shariar, M. Altaf, M. M. Hassan, S. A. AlQahtani, and A. Alsanad, "Ejection Fraction estimation using deep semantic segmentation neural network," *The Journal of Supercomputing,* vol. 79, pp. 27-50, 2023.

[21] M. Saeed, R. Muhtaseb, and M. Yaqub, "Is Contrastive Learning Suitable for Left Ventricular Segmentation in Echocardiographic Images?," *arXiv,* 2022.

[22] M. Fischer, T. Hepp, S. Gatidis, and B. Yang, "Self-supervised contrastive learning with random walks for medical image segmentation with limited annotations," *Computerized Medical Imaging and Graphics,* vol. 104, p. 102174, 2023.

[23] Z. Zhuang, P. Jin, A. N. Joseph Raj, Y. Yuan, and S. Zhuang, "Automatic segmentation of left ventricle in echocardiography based on YOLOv3 model to achieve constraint and positioning," *Computational and Mathematical Methods in Medicine,* vol. 2021, pp. 1-11, 2021.

[24] H. Zhang, A. Wahle, R. K. Johnson, T. D. Scholz, and M. Sonka, "4-D cardiac MR image analysis: left and right ventricular morphology and function," *IEEE transactions on medical imaging,* vol. 29, pp. 350-364, 2009.

[25] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Multi-label whole heart segmentation using CNNs and anatomical label configurations," in

International Workshop on Statistical Atlases and Computational Models of the Heart, 2017, pp. 190-198.

[26] D. F. Pace, A. V. Dalca, T. Geva, A. J. Powell, M. H. Moghari, and P. Golland, "Interactive whole-heart segmentation in congenital heart disease," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015, pp. 80-88.

[27] S. P. O'Brien, O. Ghita, and P. F. Whelan, "A novel model-based 3D ${+} $ Time left ventricular segmentation technique," *IEEE transactions on medical imaging,* vol. 30, pp. 461-474, 2010.

[28] S. Zambal, A. Schöllhuber, K. Bühler, and J. Hladuvka, "Fast and robust localization of the heart in cardiac MRI series," *Proc VISAPP,* vol. 1, pp. 341-6, 2008.

[29] L. Zhong, J.-M. Zhang, X. Zhao, R. S. Tan, and M. Wan, "Automatic localization of the left ventricle from cardiac cine magnetic resonance imaging: a new spectrum-based computer-aided tool," *PloS one,* vol. 9, p. e92382, 2014.

[30] Y. Lu, P. Radau, K. Connelly, A. Dick, and G. A. Wright, "Segmentation of left ventricle in cardiac cine MRI: An automatic image-driven method," in *Functional Imaging and Modeling of the Heart: 5th International Conference, FIMH 2009, Nice, France, June 3-5, 2009. Proceedings 5*, 2009, pp. 339-347.

[31] P. Kellman, X. Lu, M.-P. Jolly, X. Bi, R. Kroeker, M. Schmidt, P. Speier, C. Hayes, J. Guehring, and E. Mueller, "Automatic LV localization and view planning for cardiac MRI acquisition," *Journal of Cardiovascular Magnetic Resonance,* vol. 13, pp. 1-2, 2011.

[32] S. K. Zhou, B. Georgescu, X. S. Zhou, and D. Comaniciu, "Image based regression using boosting method," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, 2005, pp. 541-548.

[33] S. K. Zhou, J. Zhou, and D. Comaniciu, "A boosting regression approach to medical anatomy detection," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[35] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Medical image analysis,* vol. 45, pp. 94-107, 2018.

[36] Y. Lu, P. Radau, K. Connelly, A. Dick, and G. A. Wright, "Segmentation of left ventricle in cardiac cine MRI: An automatic image-driven method," in *Functional Imaging and Modeling of the Heart: 5th International Conference, FIMH 2009, Nice, France, June 3-5, 2009. Proceedings 5*, 2009, pp. 339-347.

[37] A. S. Saif, E. D. Wollega, and S. A. Kalevela, "Spatio-Temporal Features based Human Action Recognition using Convolutional Long Short-Term Deep Neural Network," *International Journal of Advanced Computer Science and Applications,* vol. 14, 2023.

[38] R. Chowdhury and A. S. Saif, "Efficient mathematical procedural model for brain signal improvement from human brain sensor activities," *International Journal of Image, Graphics and Signal Processing,* vol. 10, p. 46, 2018.

[39] M. H. Al Walid, D. Anisuzzaman, and A. S. Saif, "Data analysis and visualization of continental cancer situation by Twitter scraping," *International Journal of Modern Education and Computer Science,* vol. 11, p. 23, 2019.

[40] A. M. Choudhury, A. S. Saif, and M. Rahman, "Toddler Sensory-Motor Development for Object Manipulation by Analyzing Hand-Pose," in *Proceedings of the International Conference on Computing Advancements*, 2020, pp. 1-7.

[41] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble," *Information sciences,* vol. 307, pp. 39-52, 2015.

[42] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature," *Geoscientific model development,* vol. 7, pp. 1247-1250, 2014.

[43] A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing,* vol. 192, pp. 38-48, 2016.

[44] A. Rabbani, H. Gao, A. Lazarus, D. Dalton, Y. Ge, K. Mangion, C. Berry, and D. Husmeier, "Image-based estimation of the left ventricular cavity volume using deep learning and Gaussian process with cardio-mechanical applications," *Computerized Medical Imaging and Graphics,* vol. 106, p. 102203, 2023.

[45] K. Škardová, T. Hussain, M. Genet, and R. Chabiniok, "Effect of Spatial and Temporal Resolution on the Accuracy of Motion Tracking Using 2D and 3D Cine Cardiac Magnetic Resonance Imaging Data," in *International Conference on Functional Imaging and Modeling of the Heart*, 2023, pp. 235-244.

[46] D. Abdelrauof, M. Essam, and M. Elattar, "LVLNET: lightweight left ventricle localizer using encoder-decoder neural network," in *2019 Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, 2019, pp. 235-238.

[47] J. Park, D. Metaxas, and L. Axel, "Analysis of left ventricular wall motion based on volumetric deformable models and MRI-SPAMM," *Center for Human Modeling and Simulation,* p. 99, 1996.

[48] S. Zhou, A. AbdelWahab, J. L. Sapp, J. W. Warren, and B. M. Horáček, "Localization of ventricular activation origin from the 12-lead ECG: a comparison of linear regression with non-linear methods of machine learning," *Annals of biomedical engineering,* vol. 47, pp. 403-412, 2019.

Predicting Alzheimer's Progression in Mild Cognitive Impairment: Longitudinal MRI with HMMs and SVM Classifiers

Deep Himmatbhai Ajabani
Application Developer Lead, Source InfoTech Inc, Atlanta, Georgia

***Abstract*—The number of elderly people has increased due to the huge growth in human life expectancy over the past few decades. As a result, age-related illnesses and ailments have become more prevalent, including Alzheimer's Disease (AD). A notable deterioration in cognitive functions, particularly memory and thinking skills, characterizes Mild Cognitive Impairment (MCI), a condition that lies in the middle of normal aging and dementia. Therefore, MCI carries a noticeably higher chance of developing into AD and frequently serves as a prelude to dementia. However, using cutting-edge image processing and machine learning techniques, it is possible to examine and find underlying patterns in these complex diseases. By using these techniques, it is possible to separate groups, identify the causes of such separation, and create disease prediction models. Clinical trials, mostly using cross-sectional Magnetic Resonance Imaging (MRI) data, have extensively looked into the use of MRI for the early identification of AD and MCI. On the other hand, longitudinal studies follow the same subjects over an extended period, giving researchers the chance to investigate cross-sectional trends as well as the development of the disease. Three different techniques are put forth in this study for the analysis and assessment of the structural data found in longitudinal MRI scans. Without considering any other diagnostic measures, this information is used to forecast the progression of those who have been diagnosed with MCI. These techniques utilize Hidden Markov Models (HMMs), which capitalize on the advantages of Support Vector Machine (SVM) classifiers.**

Keywords—Alzheimer's disease; image processing; Magnetic Resonance Imaging; Mild Cognitive Impairment; machine learning

I. Introduction

The extraordinary organ known as the human brain is in charge of controlling every aspect of the body, including breathing, blood circulation, digestion, and digesting. Additionally, it acts as the control center for conscious functions including thinking, memory formation, thought retrieval, and decision-making while facilitating conscious behaviors like walking, talking, and visual perception. The brain is an equally fascinating phenomenon when seen from an anatomical standpoint. It is thought that it has about 100 billion neurons and a mind-boggling 100 trillion synapses, which are the connections between neurons that allow for communication. The network of blood arteries in the brain is essential to maintaining its normal operation. Surprisingly, the brain controls an astounding 20% of the body's blood flow despite making up just around 2% of the total body weight. Around 400 billion capillaries make up this complex circulatory system, which works ceaselessly to deliver oxygen, glucose, and other nutrients necessary for the survival of brain cells. The brain's large number of neurons plays a crucial role in preserving optimal function. The long lifespan of neurons, which begins during fetal development and lasts for up to a century, makes them unique. In the extremely rare case that they perish, neurons can regenerate, highlighting the significance of routine maintenance and repair. Individual differences in these alterations' scope and timing can have a significant impact on their impact levels. A diminished ability to learn new information, problems recalling memories, and increased difficulty performing tasks that were once simple to complete are common signs of aging. Importantly, these talents are not completely restricted because cognitively healthy senior people can still carry out these tasks, albeit somewhat more slowly than their younger counterparts. With 50–80% of dementia cases being caused by Alzheimer's Disease (AD), it becomes clear that AD is the most common type of dementia. The age of diagnosis or onset of the disease has a significant impact on the life expectancy of AD patients, which ranges from three to ten years. Although structural brain atrophy, pathological amyloid deposits, and metabolic alterations in the brain are thought to be related with Chronic Traumatic Encephalopathy (CTE), it is unclear whether these elements are the disease's causes or the results of its progression. Due to its effects on memory, Mild Cognitive Impairment (MCI), also known as amnestic MCI, is seen as an early stage of AD. Even though MCI is not considered a true disease, the early stages of AD are very similar to it. Although not severe enough to interfere with their daily lives, people who struggle with MCI experience memory, language, and judgment issues that are noticeable to others and distinct from usual aging symptoms.

The increased risk of acquiring AD in the future that MCI patients face compared to people with cognitively normal brains emphasizes the importance of MCI. As a result, MCI is a topic that medical professionals are quite interested in. There are significant ambiguities in the distinctions between the three phases of cognitive health - Cognitively Normal (CN), Moderate Cognitive Impairment (MCI), and AD - and no clear-cut standards for determining an individual's stage are in place. Nevertheless, decades of study have produced a range of approaches intended to assess the condition of brain health such as [1]–[3]. Science and medicine have placed a lot of emphasis on the study of the brain and its anomalies [4].

However, there has been a barrier to non-invasively studying it for a very long time. Even today, the only time an exact diagnosis for AD can be made is post-mortem, during the autopsy, when amyloid plaques produced in the brain and other indicators of brain degeneration can be studied by a doctor. Early identification and detection of AD and MCI are crucial because they can help patients and their families get ready for illness and start treatment as soon as feasible. In exchange, this can provide patients the chance to take part in clinical trials where the most recent medicines can be used, and they can generally manage the illness better [5]. Scientists can simultaneously study the early phases of AD and MCI to understand the disease's origin, which could result in better techniques of therapy or prevention. According to estimates, MCI patients get AD at a rate of 10–15% while cognitively healthy people develop dementia at a rate of 1%–2%. With the use of Magnetic Resonance Imaging (MRI), researchers may now conduct non-invasive in vivo examinations of the human body. This means that the brain can be monitored and evaluated to establish a baseline of what it should resemble at various stages of the disease's course or even in cognitively normal brains. It is now possible to study the brain and the changes that take place because of either the normal aging process or particular disorders thanks to the combination of computer science and machine learning.

This study's goal is to examine and interpret the structural data obtained from longitudinal brain MRI images. To investigate the gradient of anatomical and morphological changes occurring in the brain as MCI progresses to AD. Despite the widespread use of MRI scans in this sector, the goal is to forecast the possible development of AD using only this information, without the addition of other biomarkers or clinical and cognitive assessments. We use a longitudinal series of MRI scans from different people who have been given different diagnoses (CN, MCI, AD) and who are transitioning to different diagnoses (CN, MCI, AD). The goal is to avoid the possibility for human judgment errors that can occur in complicated and time-consuming processes by just using the data derived from structural (volumetric) brain changes. Given that individuals often seek medical advice after the onset of symptoms and that the diagnostic process needs some time to complete, this technique enables a quicker forecast of the condition. This study also aims to evaluate and investigate the accuracy of longitudinal MRI scans in foretelling the transition from MCI to AD in this domain. The longitudinal MRI scans are viewed as a series of observations, after which Hidden Markov Models (HMMs) [6]–[10] are used for modeling. Then, either the HMMs alone or a Support Vector Machine (SVM) [11]–[13] classifier that has been trained using the data that the HMMs used to represent the data are used to make the predictions. It's crucial to understand that this study does not try to improve, extend, or implement any one predefined technique. This technique is unusual because it uses longitudinal MRI images that are obtained one year apart and uses only the structural data that was derived from those scans. As a result, a direct comparison between the performance and results of the trials carried out for this study and the most recent findings is not possible.

The paper is as follows: In Section II we will see the related works. In Section III, an empirical study has been presented consisting of dataset description and evaluation metrics. In Section IV, the proposed models have been discussed. In Section V, the experimental results and analysis have been done. In Section VI, the thought of the paper has been presented and we conclude the paper in Section VI with some conclusions and future works.

II. Related Works

A quick overview of current research on AD and longitudinal data in the domains of computer science and machine learning is provided in this section.

A. Brain

A substantial amount of research has been focused on identifying and extracting the elements from an MRI scan that are the best diagnostic predictors in order to facilitate additional diagnosis [14]–[16]. The features that are most frequently used involve the assessment of both grey and white matter volumes [17], either over the whole brain or in particular areas such the frontal, temporal, parietal, and hippocampal cortex [18]. Furthermore, cortical thickness is a common characteristic [16], as are CSF density maps [19], [20]. Manually extracting and choosing characteristics from MRI scans is a very difficult and time-consuming process. When some feature parameters change and the feature-extraction process needs to be repeated, it often leads to the possibility of inaccurate data or complexity in re-extracting features. As a result, several tools have been created, such as FreeSurfer[1], FSL[2], and SPM[3], which let scientists and medical experts handle and interpret MRI scans in different ways and accomplish accurate feature extraction. With little to no human oversight, these technologies can carry out extraction and selection tasks.

B. Alzheimer Disease

Considerable advancements have also been achieved in the effort to control and make use of these attributes. Using different classification techniques, it is possible to distinguish between a brain that is cognitively normal and one that is impaired, either by AD or MCI. MRIs and f-MRIs are commonly used in this type of research to get anatomical and physiological brain features, which are then used to determine any pathological or normal changes occurring in the brain. These include the assessment of cortical thickness and the density of cerebrospinal fluid, as previously indicated, as well as volumetric measures of several brain areas, such as the cingulate cortex, hippocampus, and parahippocampal gyrus. These assessments make it easier to identify anomalies in the brain and offer vital information about whether dementia of any kind is present. Biomarker characteristics taken from MRI and Positron Emission Tomography (PET) scans are used in a study focused on the classification of AD and MCI [37] to enable an SVM classifier to differentiate between CN and MCI or CN and AD patients. For MCI and AD classification, the systems achieve 76.4% and 93.2% accuracy, respectively.

[1] https://surfer.nmr.mgh.harvard.edu/
[2] https://http//fsl.fmrib.ox.ac.uk/
[3] https://www.fil.ion.ucl.ac.uk/spm/

Brain biomarkers and Region-of-Interest (ROI)-based morphological parameters, such as cortical thickness values, volumes of the cerebral cortical grey matter, and cortical-associated white matter, are employed in a related investigation [16]. To create characteristics and detect abnormality patterns, this study presents the idea of correlated abnormalities, which is achieved by associating different ROIs with one another. Additionally, CN and MCI, CN and AD, and MCI and AD are separated using an SVM classifier that has been trained; the corresponding classification accuracies are 83.75%, 92.35%, and 79.24%. An Orthogonal Projection to Latent Structures (OPLS) is another technique that has been developed [21]. This technique, which was initially created for the modeling of complicated data, combines the concepts of Orthogonal Signal Correction (OSC) [22], [23] with Partial Least Squares (PLS) regression. The methodology is predicated on the idea that the observations are produced by latent variables. Nevertheless, systematic differences in the independent variables unrelated to the class labels seem to have a negative impact on it. As a result, the OPLS approach was created to deal with this problem. When separating AD from CN individuals, the OPLS classifier achieves a sensitivity of 86.1% and a specificity of 90.5%. Though they produce less-than-ideal outcomes, less-common alternatives include decision trees, Artificial Neural Networks (ANN) [24], [25], and other techniques for regression or classification. These techniques use cross-sectional MRI images and focus exclusively on the brain structure information found in the present scan to identify or forecast the disease.

C. Longitudinal Data

Even while the analysis of merely cross-sectional MRI scans has shown encouraging and effective results, it is limited in its ability to provide information about a single point in a disease's progression or the overall health of the brain. It is unable to reveal information about changes that occur over time, recognize patterns, or create connections with various circumstances. The research community has been highly interested in longitudinal studies as a result, which involve a series of data, like brain MRI scans, that are taken at regular intervals (e.g., every six months or annually). Some studies that handle longitudinal data have been centered around f-MRI scans [26], [27], focusing on the responses that particular brain areas display. Regression techniques, namely linear and modified least squares models, were primarily employed in these investigations. It is important to remember, though, that these studies cannot be classified as using longitudinal MRI scans in the sense that this study intends, since f-MRI captures brain activity over seconds, while the longitudinal data we are attempting to use investigates changes in the brain occurring over much longer time frames. The use of HMMs was first implemented to try to identify mild Alzheimer's disease, or early dementia, in older people [18]. In this work, characteristics taken from a series of MRI scan slices are combined into a time series, which is subsequently, subjected to HMM analysis and classification. With accuracy reaching up to 97.8% in some tests, the suggested strategy shows great promise in the early identification of dementia. However, the main goal of this study is to identify AD based on a single brain snapshot; it does not attempt to anticipate or explore the disease's progression across a number of years, and the data is still not truly longitudinal. Similarly, HMMs are used to predict the age of people who do not have dementia in a different study [20], but longitudinal data is not used in this instance. Prediction inaccuracy on average is as low as 2.57 years. The study in [28] introduces the use of longitudinal MRI scans to investigate changes and correlations between nine-year scans of cognitively normal and demented brains. By utilizing 9-year longitudinal MRI scans, scientists examine the data obtained from individual scans, opening a promising field with enormous potential for brain study. Even though we use different characteristics and datasets, this study is really important to our work since it shows how much information longitudinal MRI scans can provide.

III. Empirical Study

A. Dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI) provided the dataset utilized in this study. The National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), a few private pharmaceutical companies, non-profit organizations like the Alzheimer's Association (AA) and the Institute for the Study of Aging (ISA), and other organizations are among the sponsors of the ADNI research initiative, which was started in 2003. It functions in partnership with the National Institutes of Health (NIH) [29]. The main goal of ADNI is to collect and make use of longitudinal data from people who have been diagnosed with CN, MCI, or AD. Whether serial MRI scans, PET imaging, other biological markers, clinical and neuropsychological evaluations, and other data can be combined to track and characterize the development of MCI and early AD is the purpose of this study. Moreover, ADNI seeks to offer a freely available database of clinical and imaging information that clarifies changes over time in brain metabolism and structure, cognitive performance, and biomarkers in CN, MCI, and AD patients. More than 50 research facilities in the United States and Canada provide ADNI with subjects. A longitudinal MRI scan dataset containing 631 individuals was made available for this study. After their initial MRI scans, 192 of these people were diagnosed with CN, 309 as MCI, and 130 as AD. Consequently, 189 were classified as CN, 202 as MCI, and 240 as AD at the time of their most recent scans. Every person had one to three follow-up scans, spaced a year apart, with a variable number of follow-ups performed. A total of 1913 MRI scans, including 1.5T sagittal 3D T1-weighted MPRAGE MRI scans, were included in the dataset. The Freesurfer pipeline, an open source set of tools for the thorough and automated examination of important aspects of the human brain, was used for the preparation of these MRI data. The analysis included mapping of cortical grey matter thickness, estimation of architectonic boundaries from in vivo data, segmentation of hippocampal subfields, volumetric segmentation of most macroscopically visible brain structures, and several other functions. It also included inter-subject alignment based on cortical folding patterns. Given that manual study of such a vast dataset would require a lot of labor and time, automation processing was essential.

Consequently, each MRI scan yielded 55 MRI-derived regional measures, comprising 21 subcortical volumes and 34 cortical thickness values.

B. Evaluation Metrics

In this section, we employ a set of crucial metrics to meticulously evaluate the efficacy and precision of the classification and prediction techniques developed throughout this investigation. These metrics play a pivotal role in gauging the performance of the models, providing a comprehensive understanding of their capabilities. True Positives (TP) measure the subjects accurately identified as having AD, while True Negatives (TN) count those correctly classified as CN or having MCI. On the flip side, False Positives (FP) represents instances where subjects are incorrectly classified as having AD, and False Negatives (FN) denote subjects wrongly classified as CN or MCI when they indeed have AD. Sensitivity, or the True Positive Rate (TPR), showcases the proportion of TP samples (AD) correctly identified, expressed as $TPR = TP\,/\,(TP + FN)$. Specificity, or the True Negative Rate (TNR), quantifies the proportion of TN samples (CN/MCI) correctly classified, calculated as $TNR = TN\,/\,(TN + FP)$. Precision ($Positive\ Predictive\ Value - PPV$) signifies the accuracy of positive classifications (AD) among all positive predictions, defined by $PPV = TP\,/\,(TP + FP)$. The F1 Score, also used combines precision and sensitivity through their harmonic mean, offering a balanced assessment of the model's performance: $F1 = 2 * (PPV * TPR)\,/\,(PPV + TPR)$. Notably, we calculate the harmonic mean of specificity (TNR) and sensitivity (TPR) for a comprehensive evaluation. The Receiver Operating Characteristic (ROC) Curve provides a graphical overview of the model's performance across varying parameters. It plots sensitivity against 1 - specificity, with a superior model closer to the upper-left corner. The Area Under the Curve (AUC) quantifies the overall model performance. The Diagnostic Odds Ratio (DOR), an essential metric in medical research, gauges the odds of a positive test result when the disease is present compared to when it's absent. Calculated as $DOR = (Sensitivity * Specificity)\,/\,[(1 - Sensitivity) * (1 - Specificity)]$, higher DOR values signify better discriminatory test performance, ranging from 0 to infinity. Collectively, these metrics form a robust framework for the precise evaluation of the developed models, enabling a comprehensive assessment of their classification and prediction capabilities.

IV. Proposed Models

The techniques employed in this work are based on HMMs. The selection of HMMs was based on their innate capacity to efficiently interpret sequential data. Their architecture is a good representation of markov chains since it includes hidden states and their emissions, which maps to data that can be observed. Although HMMs are mainly used for markov chains, they are also widely used to capture sequential relationships in time-sequential data, like speech processing. They are a useful tool in our situation for processing the longitudinal MRI scans as observations and identifying the relationships between them, with an emphasis on the markov chain, a hidden structure. Three different techniques are presented in this study, each expanding on the preceding one. These techniques will be covered in detail and with thorough explanations in the sections that follow. To maintain clarity, we will now outline the data partitioning and utilization process that will be used in the upcoming sections. As was previously mentioned, the dataset consists of an assortment of MRI images from different people. A series of scans are available for each participant, including an initial cross-sectional scan and one to three follow-up scans. There are two basic ways to partition the data. The initial technique focuses on the diagnosis made from the first cross-sectional scan, which is known as the "subject-initial-group". The participants are classified as having MCI, being CN, or having been diagnosed with AD. No follow-up diagnoses are considered in this category; only the baseline diagnosis is taken into account. The "subject-end-group", which is the last follow-up scan diagnostic, is used to categorize individuals in the second technique. Similar to the first technique, this grouping yields the same diagnoses/categories as the subject-initial-group (CN, MCI, & AD) and only considers the diagnosis obtained from the most recent follow-up scan. Reconfiguring this data separation makes more sense in the context of this study. Although their labels have changed, the subject-initial-group and subject-end-group remain the same. The CN and AD groups are joined in the subject-initial group to produce two alternative categories: CN/AD or MCI. This modification makes more sense because the main goal is to investigate the development of the MCI subject-initial group, which is a high-risk group. Our goal is to ascertain whether MCI will progress to AD. As a result, the subject-end group falls into one of two categories: CN/MCI or AD. The training and testing sets for our models are defined by the subject-initial group; the training set is CN/AD, and the testing set is MCI. The particular interest in MCI patients and the investigation of their possible long-term progression are the driving forces behind this tactic. The wide range of cognitive impairments associated with MCI makes it a particularly important group in the field of medical research because it can progress into several disorders, including AD. The main goal is to assess how well our systems can anticipate outcomes for this population. The attempt to see how well an HMM can extract basic and generic structural changes that indicate progression toward AD or CN/MCI (conversion to CN or stability) is the rationale behind using non-MCI participants in the training set. Next, we evaluate the applicability of these derived features to the MCI group (a subset of the MCI group is used for training during experimentation to evaluate its effect on the overall performance of all techniques).

A. Technique 1: HMM Classification

In the first technique, HMMs are only used to evaluate how well they can extract and represent temporal structural changes in the brain throughout normal aging or as it moves closer to AD. The next step is to find out if these alterations are like those shown by a brain that has been diagnosed with motor cortex injury. HMMs are trained to maximize the probability $P(O|\lambda)$, where $O = [o_1, o_2, \ldots, o_T]$ denotes a series of observations, and λ denotes the HMM model. An observation series, represented by the letter O in our dataset, is equivalent to a subject's longitudinal MRI scan sequence. The volumetric data retrieved from each scan, represented by the vector o_t, has a size of [1 × 55] for each observation [2, 4]. First, the [1 × 55] sized vectors are aggregated to create observation sequences for each subject ($[n \times 55], n \in [2, 4]$). Subsequently, the non-MCI subject-initial-group data are employed to train λ_{AD} and $\lambda_{CN/MCI}$, two HMMs. Only the observations from individuals in the AD and CN/MCI subject-end groups are considered for each HMM. Following the effective training of these two HMMs, testing is conducted on the MCI subject-initial-group. Two probabilities, $P_{AD}(O_i|\lambda_{AD})\ and\ P_{CN/MCI}(O_i|\lambda_{CN/MCI})$, are calculated for every observation sequence using the forward algorithm. These probabilities show how likely it is that the matching HMM might produce each sequence. The sequences are predicted based on this probability.

$$y_i = \begin{cases} AD,\ if\ P_{AD}(O_i|\lambda_{AD}) \geq P_{\frac{CN}{MCI}}(O_i|\lambda_{\frac{CN}{MCI}}) \\ CN/MCI,\ if\ P_{AD}(O_i|\lambda_{AD}) < P_{CN/MCI}(O_i|\lambda_{CN/MCI}) \end{cases} \quad (1)$$

It is crucial to stress that we do not assign any semantics to the states in the HMMs. When using an HMM conventionally, the number of states is usually selected so that each state, or combination of states, corresponds to a unique "logical state" of the underlying process. Nevertheless, modeling the sequences in this way is not practical in our scenario because of the sparse nature of the scans. From here on, we think of the states as an independent variable in the system that we can work with and examine to see how it affects the system's overall performance.

B. Technique 2: HMM Modelling SVM Classification

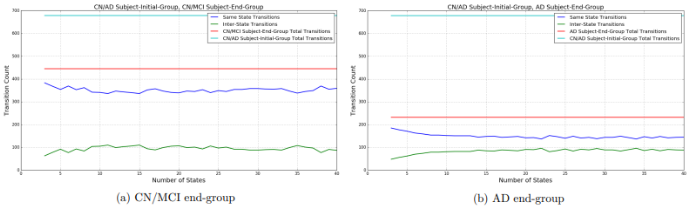
There is a premise that there is room for improvement even when the initial technique has produced good results. The data has shown that it contains useful information, and the HMM's ability to extract and represent this information has been validated. The investigation of whether this data may be organized in a way that makes it compatible with alternative models and techniques is therefore of particular interest. There is also some interest in the possibility of improving performance by adding a strong classifier. Although the HMM's states were not given explicit meanings in the previous discussion, it is agreed that the transition matrix shows that the HMM may implicitly assign certain meanings to its states after training. HMMs regard their states as markov chains by definition. As a result, the data structure for the observation variable (time) is examined during the training phase. Patterns or anomalies that appear repeatedly in the observation sequences are identified, and the start, transition, and emission probabilities are set up to correspond with these patterns. To shed further light on the reasoning behind this approach, look at the following example: Let's say the goal is to create a model of adults' everyday activities using observations made at predetermined times of the day. These findings differentiate between those who are employed and those who are not. Because of this, these two groups' observation sequences are different from one another, reflecting their different lifestyles. After being trained in this scenario, an HMM adjusts its probability to construct state markov chains that produce observation sequences that match each adult's lifestyle. The transition matrix might have been initialized to direct the HMM in giving the states particular meanings to predefine state transitions. State definitions for "working", "commuting to/from work", "sleeping", and "resting" may have been applied in the preceding case. The HMM would organize the state markov chains in a way that is easier to understand by initializing higher probabilities for state transitions like "commuting to/from work to working", "working to commuting to/from work", and "commuting to/from work to eating", and lower probabilities for transitions like "working to sleeping" and "sleeping to working". It is important to remember that while this alignment of states to actions could make sense more naturally to humans, the HMM itself may not necessarily gain from it. The HMM may nevertheless arrange the markov chains in a way that makes sense for the data's behavior even in situations where states and actions aren't directly connected. This is true even when it's not immediately obvious to human observers. The same action may even be assigned to several states by the HMM. Either way, the general organization of the states and how they behave is tailored to the features of the training set. The states of our HMMs may theoretically represent the three cognitive states of the participants (CN, MCI, and AD), much like the example given. It could have been possible to initialize the HMMs in a way that directed them to ascribe states based on the predetermined cognitive conditions right from the start. Nevertheless, we have chosen not to initialize the HMMs because of the characteristics of the data and the inherent unpredictability of these circumstances. Rather, we let them independently determine the state structure without any prior knowledge. The phrase "nature of data" refers to a range of characteristics that are taken out of every MRI scan, including the corresponding diagnosis, which is prone to inaccuracy. The three cognitive states are also rather inclusive. In particular, the MCI condition has a broad range of severity fluctuations, as was explained in previous sections; this feature also applies to the AD condition. As a result, it makes sense to believe that there are intermediate states between the states that exactly match the predetermined conditions. Because of this, we do not initialize the HMM's matrices; instead, we allow the training process to define the probabilities related to the state and emission structures. We specifically want to use this characteristic of HMM modeling in this technique. We hypothesize that important information contained in the state sequences corresponding to our observation sequences can be used as features for an alternative model or classifier. Our goal is to produce state sequences, or features, by extending the logic and foundation set by the prior technique. These features will then be used to train an SVM classifier. As previously mentioned, the procedure creates observation sequences ($O = [o_1, o_2, ..., o_T]$) for each subject. Then, using the CN/AD subject-initial-group, an HMM is trained. Unlike the prior technique, which trained multiple HMMs depending on the subject-end-group for each observation, this technique simply trains one HMM. Our goal in making this decision is to investigate the inherent capacity of the HMM to define and model discriminative information on its own, as well as to extract more generic characteristics from the data. Following training, all observation sequences (including the subject-initial-groups CN/AD and MCI) have state sequences produced by the HMM λ. These state sequences function as characteristics for the technique's next step. An SVM classifier is trained using feature sequences from the CN/AD subject-initial-group. After being trained for binary classification, the SVM divides the data into two categories: AD and CN/MCI, depending on which subject-end group belongs to which sequence. After training, the efficacy of the SVM is determined by analyzing its ability to categorize the MCI subject-initial-group sequences into the two designated classes (AD or CN/MCI).

C. Technique 3: HMM Modelling SVM Classification 2

Due to the intrinsic properties of the data, the state sequences produced by the HMM show a significant amount of volatility and fluctuation, especially when the number of states rises. Additionally, the original state sequences are features, but they invariably have varying durations based on how long the observation sequence was. This unpredictability makes the final prediction more difficult to make and adds instability to the classification process. After producing state sequences for different state counts, it is clear that, in the CN/AD and MCI datasets, the state sequences for the CN/MCI subject-end-group exhibit a more consistent pattern than those from the AD subject-end-group. There are few state transitions in the case of CN/MCI participants. This difference between the more stable conduct of CN/MCI subjects and the more erratic behavior of AD subjects may be due to structural similarities in the brain throughout normal and pathological aging. The brain regions under study typically undergo constant changes during the ordinary aging process, frequently at a slow and steady rate (e.g., a specific brain area steadily decreases in volume over the years). But because aberrant aging is characterized by unpredictable aging, these changes become more sudden and difficult to monitor precisely. There is clearly a substantial correlation even though it is still unknown if the irregular changes are the result of atypical aging or the cause of it. It may even be a combination of the two. We are mostly interested in transitions that either keep the present state or change it in the analysis of state transitions:

$$s_t \longrightarrow s_{t+1}: \begin{cases} same - state\ transition,\ if\ s_t \equiv s_{t+1} \\ inner - state\ transition,\ if\ s_t \not\equiv s_{t+1} \end{cases} \tag{2}$$

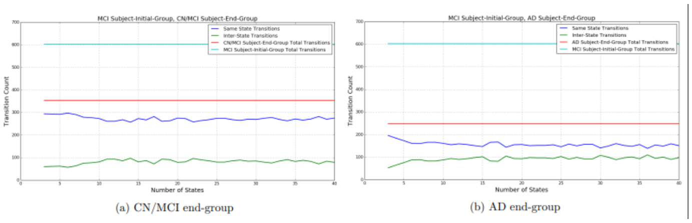
As such, we track and log the total number of transitions as well as the intra-state (same-state) and inter-state transitions that occur inside each group and end-group. The above Eq. (2) is used to determine the percentages of same-state and inter-state transitions that take place inside the AD subject-end-group of the CN/AD individuals. While 75% of the transitions are inter-state, 25% are same-state transitions. Similar computations can be made for every group and end-group to find interesting and possibly useful differences that the HMMs have brought to light. Using HMMs with different numbers of states, Fig. 1 and Fig. 2 show the counts of same-state and inter-state transitions for various groups. As a reference frame, the overall transition counts for the relevant groups are also displayed. Because they are correlated with sequence lengths—that is, the number of follow-ups scans the participants receive—rather than the structural analysis of the HMMs, these total transitions stay constant across the graphs. Interestingly, the numbers show that in the CN/MCI subject-end-groups, whether they are the CN/AD or MCI initial-groups, the number of same-state transitions is much larger (about three to four times) than the number of inter-state transitions. The AD subject-end-groups, on the other hand, show a less noticeable difference (about 1.5 to 2 times). Figs, which show the percentages of same-state and inter-state transitions compared to all transitions, support this conclusion. The Figures also compute these percentages' mean and variation across a growing number of HMM states. The information in the table highlights the fact that roughly 22% of CN/MCI end-group transitions are inter-state and 78% of them are same-state. As opposed to the CN/MCI groups, the AD end-group generates sequences where roughly 63-64% of the transitions are same-state, supporting our initial claim that the AD end-group's sequences follow a more regular pattern.



(a) CN/MCI end-group

(b) AD end-group

Fig. 1. Number of state transitions that the CN/AD group of HMMs trained with a growing number of states experienced.

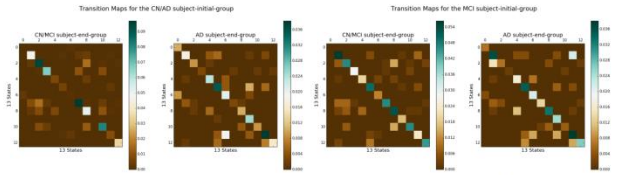


(a) CN/MCI end-group

(b) AD end-group

Fig. 2. Number of state transitions that the MCI group of HMMs trained with a growing number of states experienced.

Now, we want to take use of this property and remove the variability caused by the length of the generated features. As shown in Fig. 3, we create transition frequency maps to accomplish this. These maps are represented as $[N \times N]$ matrices, where N is the total number of HMM states. As a counter, each element in the matrix, a_{ij}, counts the number of transitions from state *i* to state *j*. Interestingly, the elements of the matrix diagonal represent inter-state transitions, whereas the components of the same-state diagonal correspond to same-state transitions. Reiterating that we treat the number of HMM states as a variable that can be adjusted to investigate its effect on system performance is crucial. Consequently, Fig. 3—13 states are just meant to serve as an illustration. These matrices are used as feature vectors for the subjects after being serialized in a row-wise manner. This technique avoids practical issues by releasing our features from the temporal element, which is intrinsic to their structure but has no bearing on their length. It is clear from the initial matrices that the feature vectors are extremely sparse, with few non-zero elements that frequently take values in the range of $a_{ij} \in [1, 2, 3]$. The feature vectors are primarily composed of zeros. As such, the non-zero components' placements are more significant than their exact values. This greatly reduces the work for an SVM classifier in comparison to the previous technique's separation of state sequences. In particular, this issue is made simpler by the fact that it may be handled by an SVM as a spatial separation problem for 2D data. The procedural stages in this technique are like those in the prior way. First, the data is prepared, and for each subject, observation sequences ($O = [o_1, o_2, \dots, o_T]$) are created. Once more, using the CN/AD subject-initial-group, another HMM is trained. Then, as previously said, state sequences are generated for each observation series, which are then utilized to make transition maps and, ultimately, converted into feature vectors. The SVM classifier is then trained using these feature vectors, with an emphasis on the CN/AD training set. Based on the end-groups of the patients, the classifier is trained to classify data into AD and CN/MCI. Ultimately, the MCI testing set is used to assess the classifier's performance. The main difference between Techniques 2 and 3 is the type of characteristics that are sent into the SVM classifier.



(a) Transition Frequency Maps of CN/AD subject-initial-group (b) Transition Frequency Maps of MCI subject-initial-group

Fig. 3. Mapping transition frequencies using a 13-State HMM.

V. Experimental Analysis

This section presents the results and evaluation of the experiments with comparison, contrast, and discussion of the different techniques.

A. Experimental Setup

Python, the hmmlearn toolbox, scikit-learn, and a number of machine learning techniques are used in the research. The construction of HMMs and SVM classifiers for the purpose of classifying subjects into various groups is the main goal of these investigations. Those without CN, those suffering from AD, and those with MCI are among these subjects. HMM models are made with the help of the hmmlearn toolkit. Since these models are fully coupled upon startup, it is possible to move between any states. Based on Gaussian emission distributions, the emission probabilities have a "spherical" covariance, which means that a single covariance value is applicable to every feature. The implementation and test run can affect how many states the HMMs have. We decided to configure the hyperparameters in advance and maintain consistency across numerous techniques and approaches for the SVM classifier. The kind of kernel, the kernel coefficient (γ), the penalty parameter (C), and the independent term for polynomial kernels are the hyperparameters that are being examined. These parameters include a polynomial kernel of degree 3, a penalty parameter C of 63.26, a γ value of 0.001, and an independent term of the polynomial function of 3.

The way the training data is handled in the experiments is one intriguing feature. The MCI subject group is first left out of the training process for the HMM and the SVM classifier by us. This technique is predicated on the idea that a greater variety and quantity of training data improve model performance and lower the likelihood of overfitting. However, choose to carry out more research to see if adding any MCI data to the training set can enhance the system's functionality. We use a technique often used in machine learning, called k-fold cross-validation, to assess the models' performance. To evaluate how well the models generalize their behavior to new data using this technique. There are k subsets of the data; k - 1 subsets are utilized for training, and the remaining subset is used for testing. The ultimate performance measure is calculated by averaging the evaluation metrics or errors generated in each run of this process, which is performed k times. We use a variant of cross-validation to incorporate MCI data into the training procedure. CN/AD and MCI participants are first given different training and testing sets. The MCI group is then divided into k = 3 folds, of which 2 are chosen as testing sets and 1 is combined with the training set. To significantly influence the process, this method yields about 25% of the training set as MCI individuals. Importantly, this modified cross-validation strategy is considered "semi-blind", whereas tests that are carried out without using MCI data in the training set are referred to as "blind". The SVM training procedures in Techniques 2 and 3 employ a conventional cross-validation procedure. The training data is split into k folds (with k values of 5, 7, or 10), either as CN/AD solely or as CN/AD plus one-third of MCI data. F1-scores are computed after training and testing several SVM classifiers. The testing set, which consists of MCI data, is classified by the SVM classifier with the greatest F1-score in preparation for the assessment.

B. Result

The following graphics contain the metrics of the various procedures under test. These metrics are assessed with and without the use of 5, 7, and 10-fold cross-validation on the SVM training, as well as with and without cross-validation on the training data. Furthermore, metrics are generated and provided for the participants who have undergone the maximum number of follow-up scans, which are three follow-ups.

1) Random classifier: Since the technique used in this study does not build upon an earlier approach, there are no state-of-the-art outcomes to compare with. As a result, the outcomes will be contrasted with a random classifier, whose performance threshold is set at the lowest possible value. Based on the values shown in Section III (A), the prior probabilities for the two classes (CN/MCI and AD) in the dataset can be defined as:

$$P_{CN/MCI} = \frac{Number\ of\ CN\ and\ MCI\ diagnoses\ at\ Last\ Scan}{Number\ of\ all\ subjects} = \frac{391}{631} = 0.62 \quad (3)$$

$$P_{AD} = \frac{Number\ of\ AD\ diagnoses\ at\ Last\ Scan}{Number\ of\ all\ subjects} = \frac{240}{631} = 0.38 \quad (4)$$

Any data point is given a class using a random classifier based on a predetermined probability:

$$P_{random} = \begin{cases} p_{class},\ where\ class\ \in [\frac{CN}{MCI}, AD] \\ \frac{1}{2},\ equal\ probability\ of\ assigning\ either\ class \end{cases} \quad (5)$$

In the first case, the classifier classifies each data point with a probability equal to the priors of the two classes, achieving the highest possible classification accuracy overall. In the second scenario, the minor class—in our case, AD—is marginally favored by the classifier. By increasing the system's sensitivity, this technique seeks to maximize the detection of AD at the cost of poor specificity, which increases the number of AD cases detected but also raises the possibility of FP results. In all scenarios, the recall, which gauges the proportion of accurately categorized data points in a particular class, stays random. According to our experimental findings, specificity is correlated with the memory of the CN/MCI class and sensitivity with the recall of the AD class. Thus, we would get the following results from the first random classifier: $Sensitivity = 0.38,\ Specificity = 0.62, and\ F1 = 0.4712$, while we would gain the following results from the second classifier: $Sensitivity = Specificity = F1 = 0.5$. The DOR for both iterations of the random classifier is 1. Currently, we have chosen the maximum values for sensitivity and specificity as our lower bounds to maximize performance optimization. As so, the following cutoff points are determined:
$Specificity_{min} = 0.62, F1_{min} = 0.554, and\ Sensitivity_{min} = 0.5.$

2) Technique 1: HMM Classification*:* Fig. 4, 5, 6, and 7 show the harmonic mean of the two (F1-score) and the sensitivity and specificity measures for increasing numbers of states. These graphs show that the number of states does not rise along with the performance of the system. The Fig. 4 and 6 show that, although the effect is not significant, there is a tendency for sensitivity to rise and specificity to fall while the F1-score remains stable. Striking for near and high values for both sensitivity and specificity is generally accepted as the standard technique. While reaching the highest possible level for both is desirable, these two metrics frequently show an adverse connection, even in the case of an absolute classifier. Positive and negative data points are accurately classified by a highly effective classifier with little to no FP and FN. Reduced specificity and sensitivity are the results of these FP and FN. A classifier's sensitivity will be almost perfect, but its specificity will be quite poor if it overclassifies one class, for example, classifying all data points as positive. In order to obtain a higher performance overall, it is wise to establish a balance between these criteria. It is clear from the blind experiment (see Fig. 4) that the sensitivity is higher than the allowable limit and the specificity first reaches the limit before declining as more states are used. Subjects with three follow-ups show a similar tendency (see Fig. 6), but the metrics are improved, leading to improved specificity performance and a delayed divergence between the two metrics. Both times, the F1-score greatly exceeds the upper bound. On the other hand, noticeably more stable metric graphs are produced by the semi-blind experiment (see Fig. 5). With very few exceptions, specificity usually reaches or exceeds the 0.62 threshold while sensitivity stays high. The stability of the graphs includes both the proximity of the two measures when compared and fluctuations in each metric separately. Similar behavior is shown in participants who have had three follow-ups, with higher metrics and a particularly noticeable improvement in specificity (see Fig. 7). The evolution of the DOR for the blind and semi-blind tests, with the full MCI group or just participants who had three follow-ups, is shown in Fig. 8. With all ratios remaining well over 2 (with 1 as the limit) and exhibiting little volatility, excellent results are seen in this case. When subjects receive three follow-ups, the blind experiment performs best in terms of DOR.

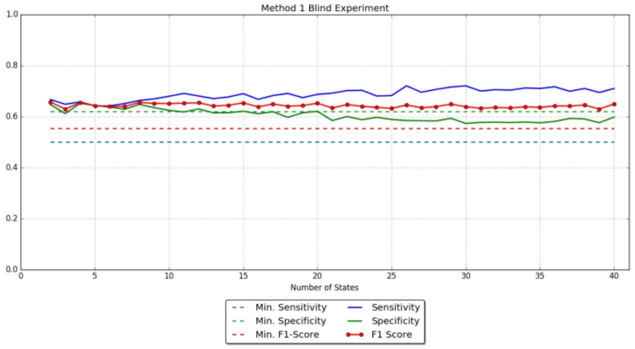


Fig. 4. Technique 1 with blind experiments: sensitivity and specificity for increasing number of HMM states (F1 score on average is 0.64).

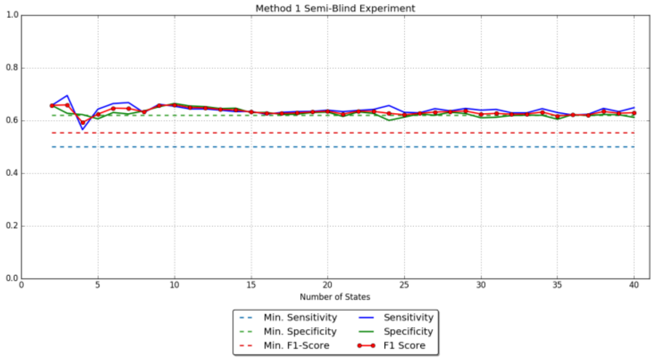


Fig. 5. Technique 1 with semi-blind experiments: sensitivity and specificity for increasing number of HMM states (F1 score on average is 0.63).

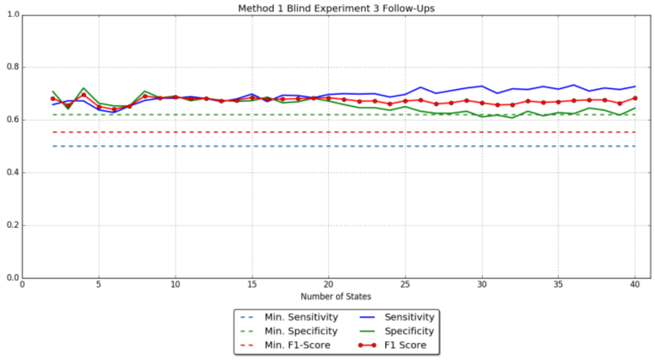


Fig. 6. Technique 1 with 3 blind experiments: sensitivity and specificity for increasing number of HMM states (F1 score on average is 0.67).

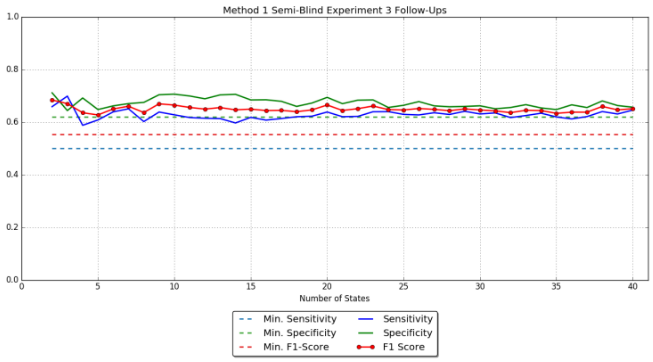


Fig. 7. Technique 1 with 3 semi-blind experiments: sensitivity and specificity for increasing number of HMM states (F1 score on average is 0.64).

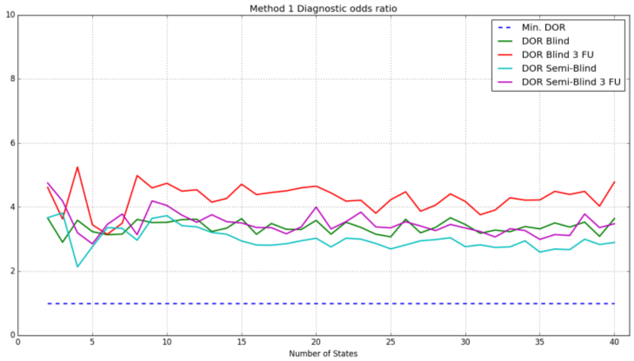


Fig. 8. DOR for the different approaches of technique 1.

3) Technique 2: HMM Modelling SVM Classification*:* The graphs take on greater interest when the second technique is examined in Fig. 9 to Fig. 20. This technique shows the F1 scores, sensitivity, and specificity for every experiment variant—Blind/Semi-Blind, all/only three follow-up scans, and each unique number of SVM folds (5, 7, & 10). Fig. 9, 10, and 11 show that the graphs for the various folds within the same type of trial are nearly identical, suggesting that SVM cross-validation has no discernible impact on system performance. Notably, the second technique shows extremely low sensitivity and very high specificity (approaching 1 for participants with three follow-up scans), which results in a low F1 score. The preceding section covered the phenomena of inverse behavior between sensitivity and specificity. After doing a thorough study of the data and examining the confusion matrices generated (see Fig. 24), it is apparent that the classifier primarily classifies most of the data as CN/MCI, which accounts for the remarkably high specificity and low sensitivity. The significant variety of the state sequences produced by the HMMs and utilized as feature vectors for the SVM makes them non-separable data, as was previously discussed. As a result, the SVM finds it difficult to identify an appropriate separating hyperplane. Upon reviewing the DORs see Fig. 21, 22, and 23) for this technique, it is evident that the system's behavior has not been affected by cross-validation for SVM, since all DOR graphs remain almost the same for varying numbers of folds. The DOR values are extremely low—much lower than those obtained using technique I. DORs in the blind tests may fall to levels less than 1. With this technique, adding MCI cases to the training set results in a marginal improvement in performance, which is mainly manifested in a smaller sensitivity/specificity divergence. Even yet, the overall outcomes are still unimpressive. Fig. 24 show two examples of confusion matrices for several experiment runs that show how many data points were categorized into each class. The confusion matrix's decimal numbers can be explained by the fact that all trials, as mentioned in Section III (A), are run ten times in order to reduce the impact of outliers. The average number of data points classified in each class over ten experiments with the same settings is effectively represented by these decimal figures. The creation of Technique 3 was spurred by the significant differences in sensitivity and specificity found in this technique.

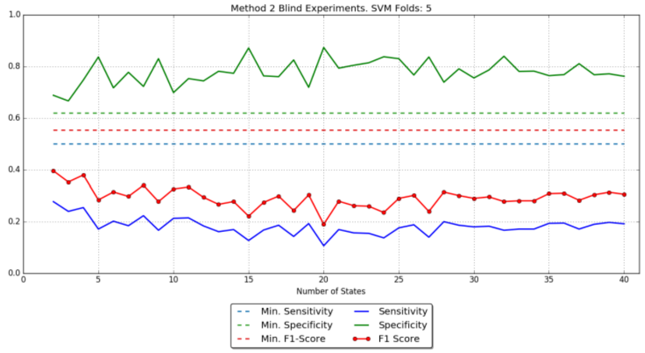


Fig. 9. Technique 2 with blind experiments: sensitivity and specificity for increasing number of HMM states with 5 SVM folds (F1 score on average is 0.29).

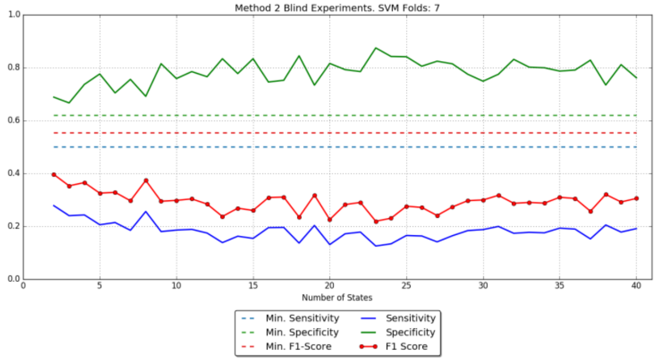


Fig. 10. Technique 2 with blind experiments: sensitivity and specificity for increasing number of HMM states with 7 SVM folds (F1 score on average is 0.29).

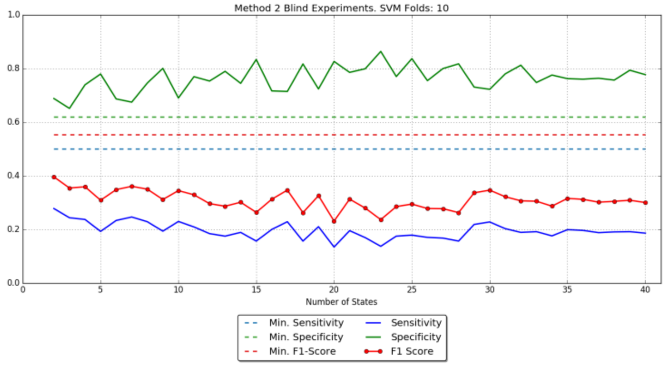


Fig. 11. Technique 2 with blind experiments: sensitivity and specificity for increasing number of HMM states with 10 SVM folds (F1 score on average is 0.30).

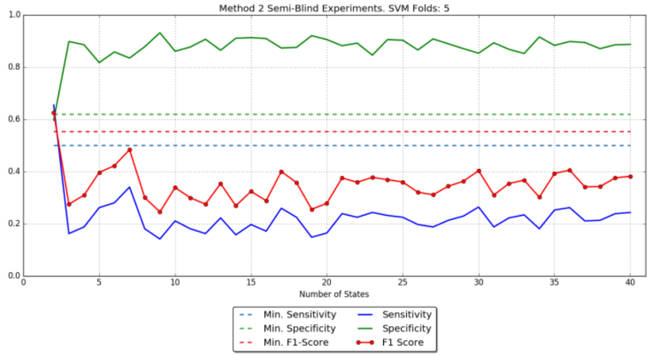


Fig. 12. Technique 2 with semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 5 SVM folds (F1 score on average is 0.34).

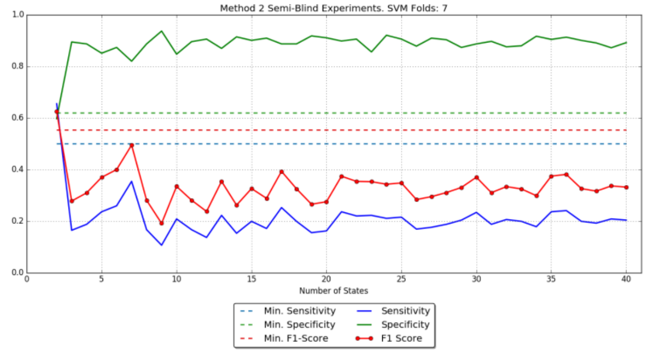


Fig. 13. Technique 2 with semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 7 SVM folds (F1 score on average is 0.33).

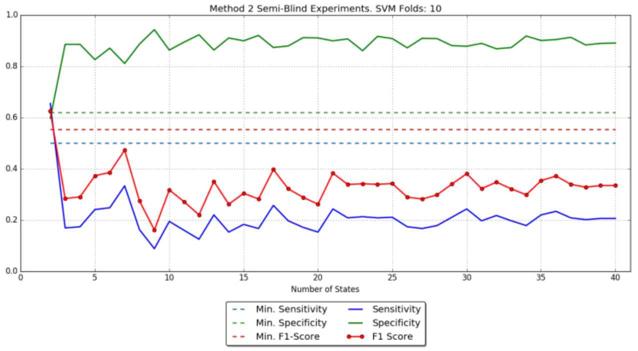


Fig. 14. Technique 2 with semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 10 SVM folds (F1 score on average is 0.32).

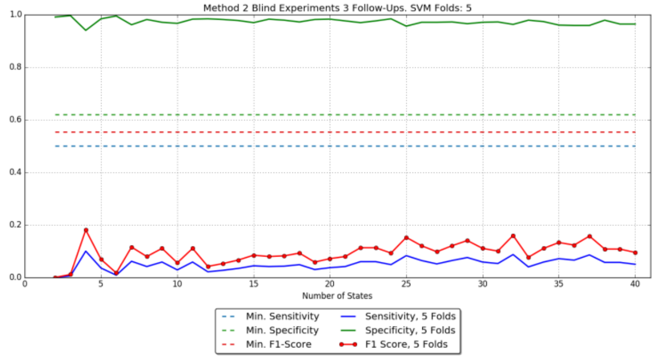


Fig. 15. Technique 2 with 3 blind experiments: sensitivity and specificity for increasing number of HMM states with five SVM folds (F1 score on average is 0.09).

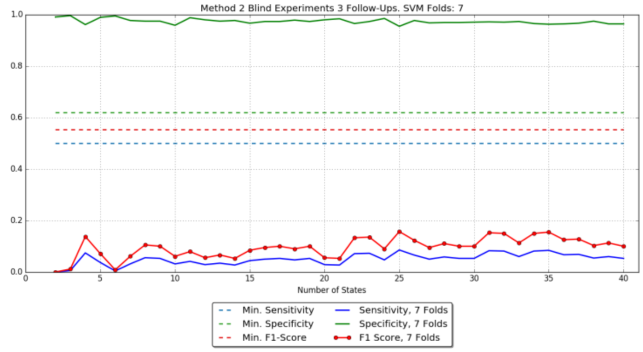


Fig. 16. Technique 2 with 3 blind experiments: sensitivity and specificity for increasing number of HMM states with 7 SVM folds (F1 score on average is 0.09).

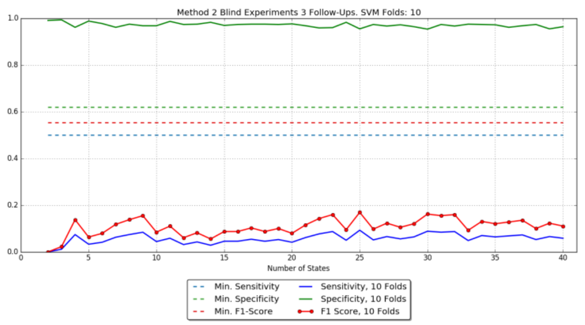


Fig. 17. Technique 2 with 3 blind experiments: sensitivity and specificity for increasing number of HMM states with 10 SVM folds (F1 score on average is 0.15).

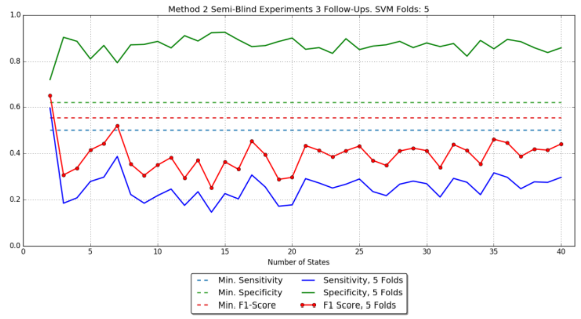


Fig. 18. Technique 2 with 3 semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 5 SVM folds (F1 score on average is 0.39).

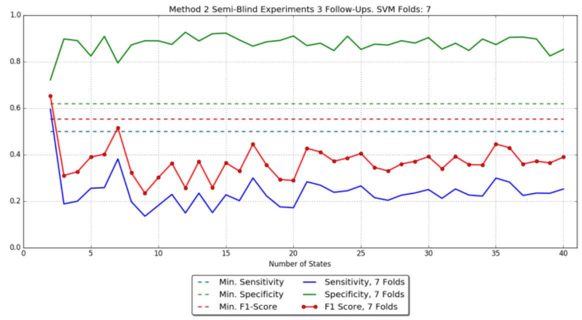


Fig. 19. Technique 2 with 3 semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 7 SVM folds (F1 score on average is 0.36).

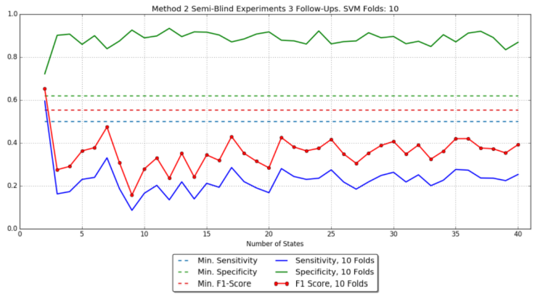


Fig. 20. Technique 2 with 3 semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 10 SVM folds (F1 score on average is 0.35).

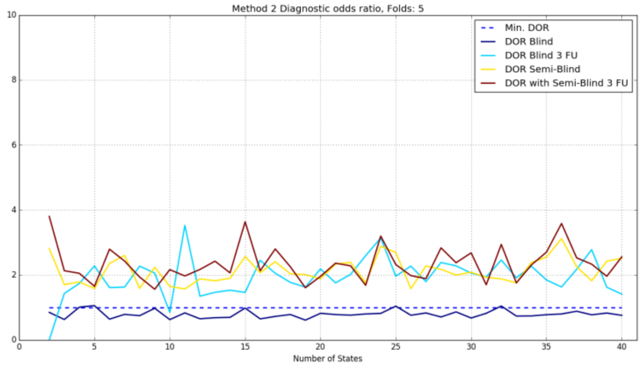


Fig. 21. DOR for the different approaches of Technique 2 with 5 folds.

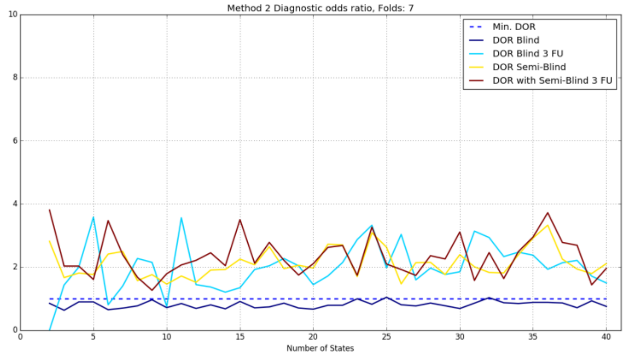


Fig. 22. DOR for the different approaches of Technique 2 with 7 folds.

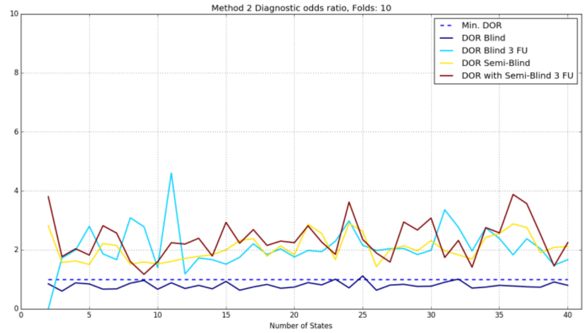
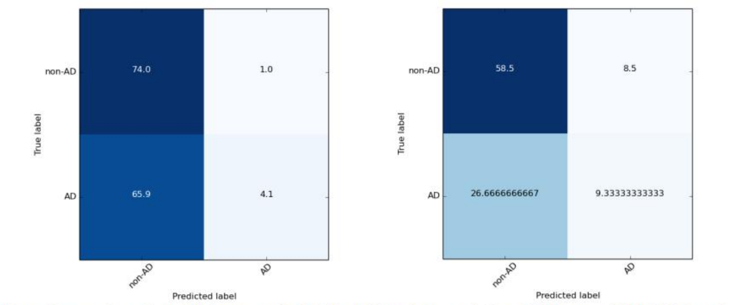


Fig. 23. DOR for the different approaches of Technique 2 with 10 folds.



(a) 11 States without Cross-Validation, 10 SVM Folds & Subjects with 3 Follow-Ups

(b) 17 States with Cross-Validation, 5 SVM Folds & with all MCI Subjects

Fig. 24. Confusion matrix for Technique 2.

4) Technique 3: HMM Modelling SVM Classification 2: In the context of the third technique, the sensitivity, specificity, and F1 score graphs for the different tests and varying numbers of SVM folds are displayed once more (see Fig. 25- Fig. 36). SVM cross-validation is often found to have minimal impact on system performance when the number of folds is changed. It is important to note that the semi-blind trials are greatly impacted by the participation of MCI participants. As demonstrated in Technique 1, the disparity between sensitivity and specificity is reduced (see Fig. 28, 29, 30, 34, 35 and 36). Not only is there no divergence here, but convergence is observed. Both metrics cross over in each of the Figures and then stay rather close after that. Fig. 31– Fig. 33 show how, during the experiment, sensitivity and specificity closely coincide with one another. Sensitivity/specificity graphs and DOR graphs are provided, much like in the previous two techniques. Specificity faces difficulties in the blind trial, falling to and remaining at the lowest threshold. But there is a noticeable improvement when compared to the second technique, suggesting that using frequency maps instead of the real state sequences makes the data easier to separate and preserves important details about the evolution of the condition. The semi-blind studies do, in fact, help to lessen the sensitivity/specificity gap, although the F1 score, and both measures show a modest reduction (see Fig. 28–30). With values far above 3.0, the DOR graphs in Fig. 37– Fig. 39 show a discernible improvement over the previous technique. When MCI participants are included in the training process, these data show a reduction that is comparable to that seen in the sensitivity/specificity graphs.

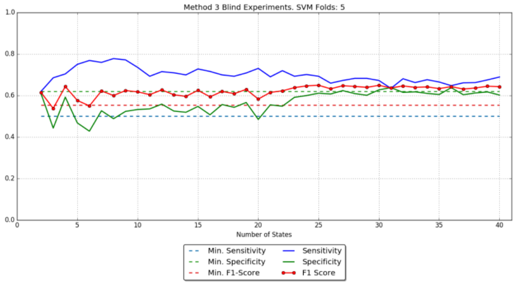


Fig. 25. Technique 3 with blind experiments: sensitivity and specificity for increasing number of HMM states with 5 SVM folds (F1 score on average is 0.62).

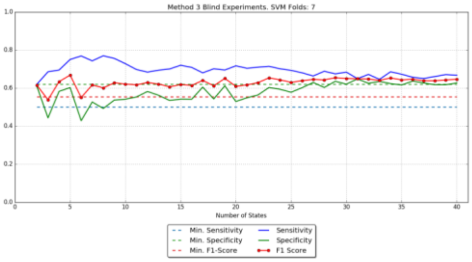


Fig. 26. Technique 3 with blind experiments: sensitivity and specificity for increasing number of HMM states with 7 SVM folds (F1 score on average is 0.62).

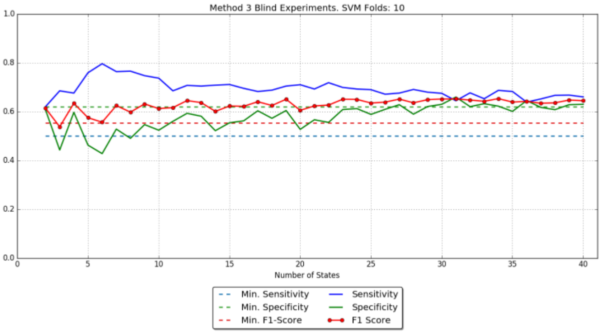


Fig. 27. Technique 3 with blind experiments: sensitivity and specificity for increasing number of HMM states with 10 SVM folds (F1 score on average is 0.62).

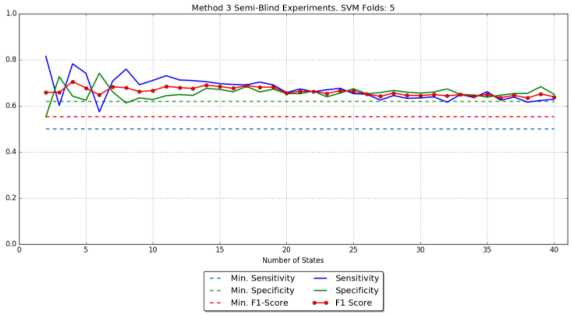


Fig. 28. Technique 3 with semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 5 SVM folds (F1 score on average is 0.66).

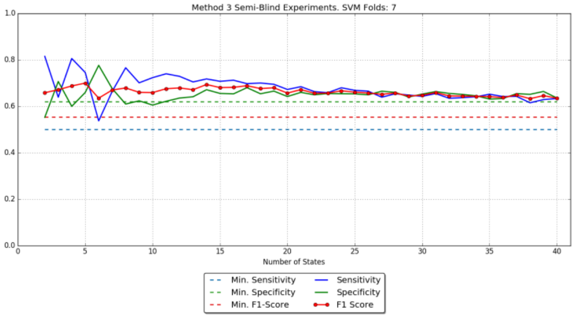


Fig. 29. Technique 3 with semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 7 SVM folds (F1 score on average is 0.66).

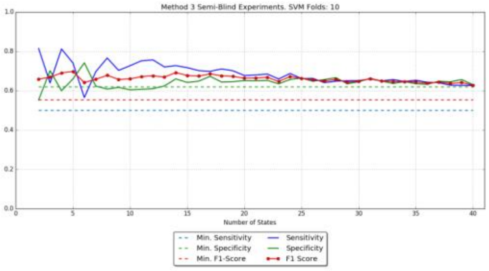


Fig. 30. Technique 3 with semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 10 SVM folds (F1 score on average is 0.66).

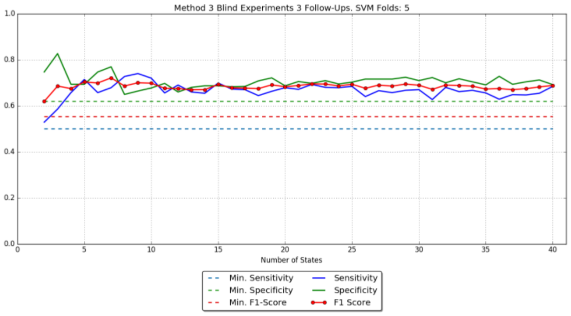


Fig. 31. Technique 3 with 3 blind experiments: sensitivity and specificity for increasing number of HMM states with 5 SVM folds (F1 score on average is 0.68).

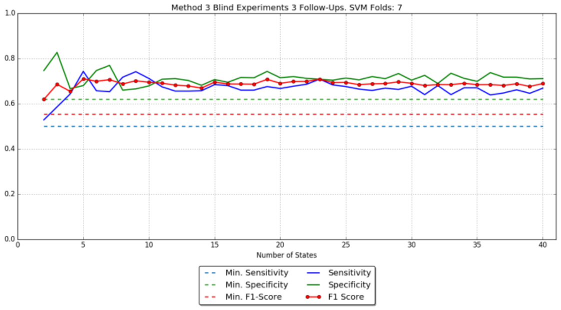


Fig. 32. Technique 3 with 3 blind experiments: sensitivity and specificity for increasing number of HMM states with 7 SVM folds (F1 score on average is 0.68).

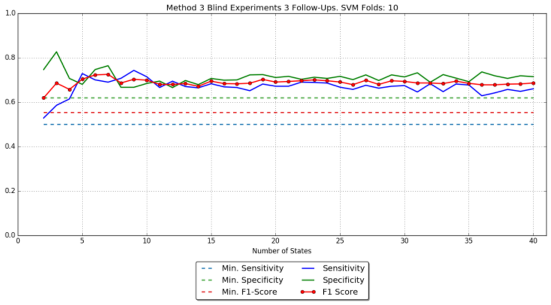


Fig. 33. Technique 3 with 3 blind experiments: sensitivity and specificity for increasing number of HMM states with 10 SVM folds (F1 score on average is 0.68).

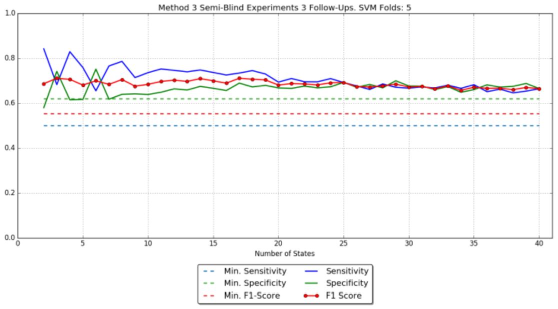


Fig. 34. Technique 3 with 3 semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 5 SVM folds (F1 score on average is 0.68).

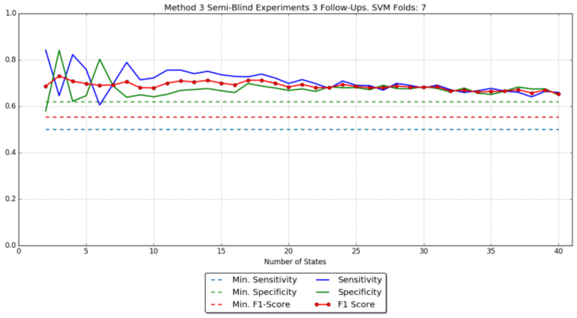


Fig. 35. Technique 3 with 3 semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 7 SVM folds (F1 score on average is 0.68).

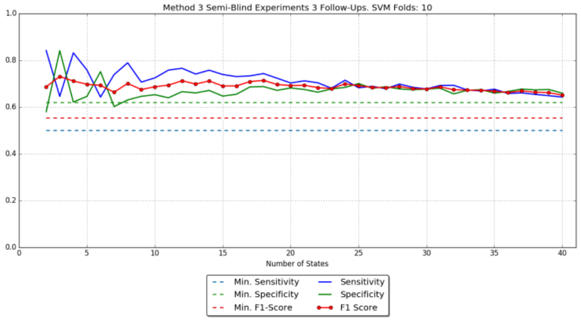


Fig. 36. Technique 3 with 3 semi-blind experiments: sensitivity and specificity for increasing number of HMM states with 10 SVM folds (F1 score on average is 0.68).

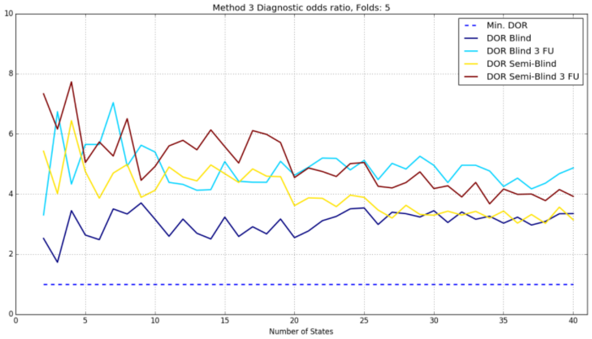


Fig. 37. DOR for the different approaches of Technique 3 with 5 folds.

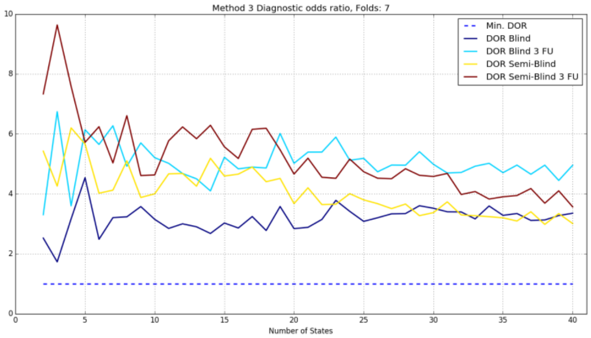


Fig. 38. DOR for the different approaches of Technique 3 with 7 folds.

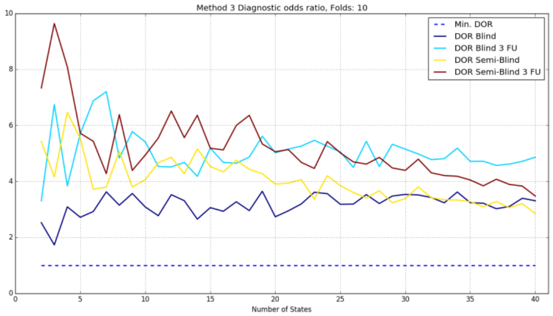
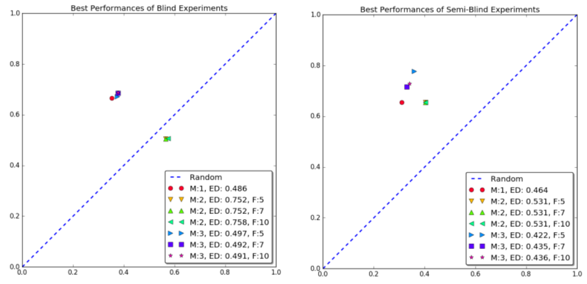


Fig. 39. DOR for the different approaches of Technique 3 with 10 folds.

VI. Discussion

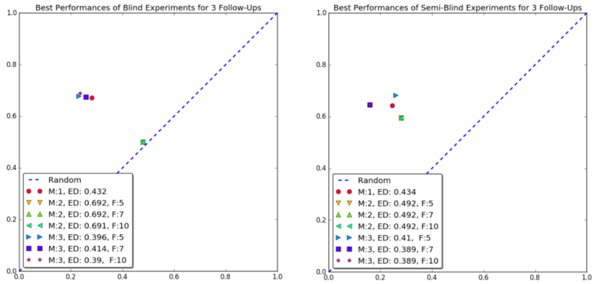
Thus far, the analyses and findings have primarily operated at a theoretical level, focusing on the techniques from a broader scientific perspective and integrating sensitivity, specificity, and DOR graphs generated across multiple states. On the other hand, specific instances of the trained and evaluated classifiers would be of interest to us if these techniques were to be applied in real life. Put another way, we want to assess the best "runs" of each technique rather than comparing results for the HMMs across various numbers of states (the major variable). To do this, we created Fig. 40 and Fig. 41, which show a ROC space filled with data points that indicate the greatest examples of each technique. For every technique, these data points represent the ideal run's TPR and 1−TNR. The various techniques are denoted by the letter M in the legend of these figures. Furthermore, we present the Euclidean distance (ED) of every point as measured from the upper-left corner; smaller EDs denote better performance. Tables I and II also provide a summary of these numerical values. The various methodologies and techniques can be more easily compared and contrasted thanks to this tabulation and visualization, which presents the performances visually. These Figures clearly show that adding MCI participants to the training set improves all tested techniques' peak points, with technique 2 showing the greatest improvement. Although Technique 2's initial performance was lower than that of a random classifier, it eventually obtains performance comparable to the other two (see Fig. 40). Furthermore, Technique 1 and 3 yield almost identical results, with Technique 3 slightly outperforming, especially for participants who receive three follow-up scans. An additional important inference from these Figures, which is further supported by the results of the separate technique, is the importance of prolonged MRI sequences. We regularly see that participants with three follow-up scans had better results than the overall results in all the presented Figures. This implies that there may be a greater likelihood of detecting AD progression, MCI progression, or perhaps conversion to CN in these individuals. It emphasizes how important it is for each person to get as many follow-up scans as possible to provide more precise projections. Moreover, selecting a different number of folds for SVM training has little effect on performance, consistent with other findings. One noteworthy observation from the data shown in Tables I and II is that although the metrics for sensitivity and specificity increase for participants who receive three follow-ups, the relative importance of the measures is inverted. When classifying all individuals, our technique shows higher sensitivity; but, when applied to participants with longer MRI sequences, it shows higher specificity. The unequal distribution of diagnosis across various numbers of follow-up scans can be used to justify this.



(a) All Methods Without Cross-Validation

(b) All Methods With Cross-Validation

Fig. 40. Best performance of all techniques.



(a) All Methods Without Cross-Validation, Subjects with 3 Follow-ups

(b) All Methods With Cross-Validation, Subjects with 3 Follow-ups

Fig. 41. Best performance of all techniques with 3 follow-ups.

TABLE I. An Overview of Each Method's Best Outcomes in Relation to the Point on the Roc Space That is Closest to the Upper Left Corner

Technique	Type	Specificity 5, 7, 10 Folds	Sensitivity 5, 7, 10 Folds	Distance 5, 7, 10 Folds	Avg. F1 5, 7, 10 Folds	Avg. DOR 5, 7, 10 Folds
1	Blind	0.646	0.665	0.486	0.643	3.381
1	Semi-Blind	0.689	0.655	0.463	0.633	2.998
2	Blind	0.434, 0.434, 0.425	0.504, 0.504, 0.505	0.752, 0.752, 0.757	0.293, 0.293, 0.31	0.794, 0.809, 0.797
2	Semi-Blind	0.596, 0.597, 0.597	0.655, 0.654, 0.654	0.530, 0.530, 0.530	0.35, 0.333, 0.329	2.135, 2.123, 2.066
3	Blind	0.626, 0.622, 0.622	0.672, 0.685, 0.686	0.496, 0.491, 0.490	0.622, 0.629, 0.628	3.054, 3.172, 3.173
3	Semi-Blind	0.641, **0.67**, 0.659	**0.776**, 0.715, 0.728	**0.422**, 0.535, 0.66	**0.662**, 0.661, 0.66	4.018, **4.04,** 4.015

Note: The point's corresponding specificity and sensitivity are displayed in the table along with its Euclidean distance from the upper left corner. The type of column is in line with the kind of experiment that was conducted using the training data. We also provide the average DOR and F1 Scores for each approach as a point of comparison.

TABLE II. An Overview of Each Method's Best Outcomes in Relation to the Point on the Roc Space That is Closest to the Upper Left Corner with three Follows Ups

Technique	**Type**	**Specificity 5, 7, 10 Folds**	**Sensitivity 5, 7, 10 Folds**	**Distance 5, 7, 10 Folds**	**Avg. F1 5, 7, 10 Folds**	**Avg. DOR 5, 7, 10 Folds**
1	Blind	0.719	0.671	0.431	0.672	4.289
1	Semi-Blind	0.753	0.642	0.434	0.649	3.495
2	Blind	0.521, 0.521, 0.522	0.5, 0.5, 0.5	0.691, 0.691, 0.691	0.095, 0.096, 0.108	1.939, 1.998, 2.089
2	Semi-Blind	0.72, 0.72, 0.72	0.595, 0.595, 0.595	0.491, 0.491, 0.491	0.391, 0.369, 0.356	2.347, 2.326, 2.357
3	Blind	0.769, 0.742, 0.764	0.677, 0.675, **0.688**	0.395, 0.414, 0.39	0.684, **0.688, 0.688**	4.855, 5.045, 5.043
3	Semi-Blind	0.74, **0.84**, **0.84**	0.682, 0.645, 0.645	0.410, **0.388, 0.388**	0.685, 0.687, 0.687	5.01, **5.177**, 5.171

VII. Conclusion and Future Works

To sum up, the main goal of this study was to create a model that could be used to predict how patients with MCI would progress based only on the examination of their longitudinal MRI scans. Another goal of this study was to use MRI data to derive useful diagnostic information without the need for further diagnostic instruments like cognitive tests. Predicting whether patients with MCI would develop AD was the third goal. Three different techniques based on HMMs were developed, one building on the other. It is clear from looking at the experimental findings that Techniques 1 and 3 have generated models that work well. These techniques have produced results that are both much higher than the preset criteria for sensitivity and higher than even the harsher predefined threshold for specificity. Particularly, technique 3 performs the best, closely followed by Technique 1. Technique 3, the best classifier, can identify 77.6% of participants who advance to AD and 64.1% of people who remain stable with MCI or return to normal cognitive status. It uses a semi-blind method with 5-fold SVM training. Furthermore, the findings highlight the structural information contained in the MRI images, which provides important information on how a patient's cognitive state is developing. Notably, the first technique relies on HMMs as the primary classifier without the requirement for a secondary classifier, using only the structural and temporal information from the scans for categorization. The findings further highlight the importance of the longitudinal MRI sequences' duration. Longer sequences consistently result in greater system performance, especially those with three follow-up scans. This emphasizes how crucial it is to get more follow-up scans for every person to improve prediction accuracy. The dataset's limitations present difficulties for future investigation. A primary concern pertains to the size of the dataset, which is determined by the expense and duration of MRI scans as well as the preprocessing done with Freesurfer. One can focus on decreasing processing time, increasing the dataset with longer sequences, and optimizing preprocessing efficiency through parallelization. A flag to denote confirmed diagnoses might also be introduced to alleviate the uncertainty around diagnoses. This would enable weighting of subjects with certain diagnoses and enable semi-supervised learning. Additional preprocessing procedures, like clustering or dimensionality reduction, to simplify and improve the feature vectors might be added in the future. It may also be worthwhile to investigate more sophisticated machine learning methods, such as Convolutional Neural Networks (ConvNets), Deep Neural Networks (DeepNets), and Recurrent Neural Networks (RNNs). Without a set input length, RNNs excel at modeling sequential and temporal data. Conversely, DeepNets and ConvNets provide high precision modeling of complicated data, which may minimize the need for intensive preprocessing. ConvNets may be able to operate directly with raw MRI scans, omitting the feature extraction stage, albeit their use may necessitate much bigger datasets. Although these methods offer fascinating directions for future study, their applicability in medicine will depend on the availability of data and developments in neural network technology.

VIII. Declarations

Funding: No funds, grants, or other support was received.

Conflict of Interest: The author declare that they have no known competing for financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability: Data will be made on reasonable request.

Code Availability: Code will be made on reasonable request.

References

[1] H. Yoo, “Genetics of Autism Spectrum Disorder: Current Status and Possible Clinical Applications,” Experimental Neurobiology, vol. 24, no. 4, pp. 257–272, Dec. 2015, doi: 10.5607/en.2015.24.4.257.

[2] K. D. Miller, M. Fidler ‐ Benaoudia, T. H. Keegan, H. S. Hipp, A. Jemal, and R. L. Siegel, “Cancer statistics for adolescents and young adults, 2020,” CA: A Cancer Journal for Clinicians, vol. 70, no. 6, pp. 443 – 459, Nov. 2020, doi: 10.3322/caac.21637.

[3] A. A. Adegun, S. Viriri, and R. O. Ogundokun, “Deep Learning Approach for Medical Image Analysis,” Computational Intelligence and Neuroscience, vol. 2021, 2021, doi: 10.1155/2021/6215281.

[4] M. M. Bronstein, J. Bruna, Y. Lecun, A. Szlam, and P. Vandergheynst, “Geometric Deep Learning: Going beyond Euclidean data,” IEEE Signal Processing Magazine, vol. 34, no. 4. pp. 18–42, 2017. doi: 10.1109/MSP.2017.2693418.

[5] J. Shaw, F. Rudzicz, T. Jamieson, and A. Goldfarb, “Artificial Intelligence and the Implementation Challenge,” J Med Internet Res 2019;21(7):e13659 https://www.jmir.org/2019/7/e13659, vol. 21, no. 7, p. e13659, Jul. 2019, doi: 10.2196/13659.

[6] N. Marwah, V. K. Singh, G. S. Kashyap, and S. Wazir, “An analysis of the robustness of UAV agriculture field coverage using multi-agent reinforcement learning,” International Journal of Information Technology (Singapore), vol. 15, no. 4, pp. 2317–2327, May 2023, doi: 10.1007/s41870-023-01264-0.

[7] S. Wazir, G. S. Kashyap, and P. Saxena, “MLOps: A Review,” Aug. 2023, Accessed: Sep. 16, 2023. [Online]. Available: https://arxiv.org/abs/2308.10908v1

[8] M. Kanojia, P. Kamani, G. S. Kashyap, S. Naz, S. Wazir, and A. Chauhan, "Alternative Agriculture Land-Use Transformation Pathways by Partial-Equilibrium Agricultural Sector Model: A Mathematical Approach," Aug. 2023, Accessed: Sep. 16, 2023. [Online]. Available: https://arxiv.org/abs/2308.11632v1

[9] H. Habib, G. S. Kashyap, N. Tabassum, and T. Nafis, "Stock Price Prediction Using Artificial Intelligence Based on LSTM– Deep Learning Model," in Artificial Intelligence & Blockchain in Cyber Physical Systems: Technologies & Applications, CRC Press, 2023, pp. 93–99. doi: 10.1201/9781003190301-6.

[10] G. S. Kashyap, D. Mahajan, O. C. Phukan, A. Kumar, A. E. I. Brownlee, and J. Gao, "From Simulations to Reality: Enhancing Multi-Robot Exploration for Urban Search and Rescue," Nov. 2023, Accessed: Dec. 03, 2023. [Online]. Available: https://arxiv.org/abs/2311.16958v1

[11] G. S. Kashyap, K. Malik, S. Wazir, and R. Khan, "Using Machine Learning to Quantify the Multimedia Risk Due to Fuzzing," Multimedia Tools and Applications, vol. 81, no. 25, pp. 36685–36698, Oct. 2022, doi: 10.1007/s11042-021-11558-9.

[12] G. S. Kashyap, A. E. I. Brownlee, O. C. Phukan, K. Malik, and S. Wazir, "Roulette-Wheel Selection-Based PSO Algorithm for Solving the Vehicle Routing Problem with Time Windows," Jun. 2023, Accessed: Jul. 04, 2023. [Online]. Available: https://arxiv.org/abs/2306.02308v1

[13] S. Wazir, G. S. Kashyap, K. Malik, and A. E. I. Brownlee, "Predicting the Infection Level of COVID-19 Virus Using Normal Distribution-Based Approximation Model and PSO," Springer, Cham, 2023, pp. 75–91. doi: 10.1007/978-3-031-33183-1_5.

[14] Y. Chen and T. D. Pham, "Sample entropy and regularity dimension in complexity analysis of cortical surface structure in early Alzheimer's disease and aging," Journal of Neuroscience Methods, vol. 215, no. 2, pp. 210–217, May 2013, doi: 10.1016/j.jneumeth.2013.03.018.

[15] S. Duchesne, A. Caroli, C. Geroldi, D. L. Collins, and G. B. Frisoni, "Relating one-year cognitive change in mild cognitive impairment to baseline MRI features," NeuroImage, vol. 47, no. 4, pp. 1363–1370, Oct. 2009, doi: 10.1016/j.neuroimage.2009.04.023.

[16] C. Y. Wee, P. T. Yap, and D. Shen, "Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns," Human Brain Mapping, vol. 34, no. 12, pp. 3411–3425, Dec. 2013, doi: 10.1002/hbm.22156.

[17] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," NeuroImage, vol. 55, no. 3, pp. 856–867, Apr. 2011, doi: 10.1016/j.neuroimage.2011.01.008.

[18] Y. Chen and T. D. Pham, "Development of a brain MRI-based hidden Markov model for dementia recognition," BioMedical Engineering Online, vol. 12, no. SUPPL 1, pp. 1–16, Dec. 2013, doi: 10.1186/1475-925X-12-S1-S2.

[19] S. M. Resnick, D. L. Pham, M. A. Kraut, A. B. Zonderman, and C. Davatzikos, "Longitudinal magnetic resonance imaging studies of older adults: A shrinking brain," Journal of Neuroscience, vol. 23, no. 8, pp. 3295–3301, Apr. 2003, doi: 10.1523/jneurosci.23-08-03295.2003.

[20] B. Wang and T. D. Pham, "MRI-based age prediction using hidden Markov models," Journal of Neuroscience Methods, vol. 199, no. 1, pp. 140–145, Jul. 2011, doi: 10.1016/j.jneumeth.2011.04.022.

[21] G. Spulber et al., "An MRI-based index to measure the severity of Alzheimer's disease-like structural pattern in subjects with mild cognitive impairment," Journal of Internal Medicine, vol. 273, no. 4, pp. 396–409, Apr. 2013, doi: 10.1111/joim.12028.

[22] J. Trygg and S. Wold, "O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter," in Journal of Chemometrics, Jan. 2003, vol. 17, no. 1, pp. 53–64. doi: 10.1002/cem.775.

[23] S. Wold, J. Trygg, A. Berglund, and H. Antti, "Some recent developments in PLS modeling," in Chemometrics and Intelligent Laboratory Systems, Oct. 2001, vol. 58, no. 2, pp. 131–150. doi: 10.1016/S0169-7439(01)00156-3.

[24] C. Aguilar et al., "Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment," Psychiatry Research - Neuroimaging, vol. 212, no. 2, pp. 89–98, May 2013, doi: 10.1016/j.pscychresns.2012.11.005.

[25] J. Escudero, J. P. Zajicek, and E. Ifeachor, "Machine Learning classification of MRI features of Alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2011, pp. 7957–7960. doi: 10.1109/IEMBS.2011.6091962.

[26] K. Katanoda, Y. Matsuda, and M. Sugishita, "A spatio-temporal regression model for the analysis of functional MRI data," NeuroImage, vol. 17, no. 3, pp. 1415–1428, Nov. 2002, doi: 10.1006/nimg.2002.1209.

[27] A. Quirós, R. M. Diez, and D. Gamerman, "Bayesian spatiotemporal model of fMRI data," NeuroImage, vol. 49, no. 1, pp. 442–456, Jan. 2010, doi: 10.1016/j.neuroimage.2009.07.047.

[28] Y. Wang, S. M. Resnick, and C. Davatzikos, "Spatio-temporal analysis of brain MRI images using hidden Markov models," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2010, vol. 6362 LNCS, no. PART 2, pp. 160–168. doi: 10.1007/978-3-642-15745-5_20.

[29] S. G. Mueller et al., "Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI)," Alzheimer's and Dementia, vol. 1, no. 1. No longer published by Elsevier, pp. 55–66, Jul. 01, 2005. doi: 10.1016/j.jalz.2005.06.003.

Integrating Social Media Data and Historical Stock Prices for Predictive Analysis: A Reinforcement Learning Approach

Mei Li[1], Ye Zhang[2*]
College of Economic and Management, Jiangsu College of Engineering and Technology, Nantong 226000, Jiangsu, China[1]
College of Economic and Management, University of Padova, Padova 35100, Veneto, Italy[2]

***Abstract*—The reliance on data collection for assessing individual behavior and actions has intensified, particularly with the proliferation of digital platforms. People often use the Internet to express their opinions and experiences about various products and services on social media and personal websites. Concurrently, the stock market, a key driver of commercial and industrial growth, has seen a surge in research focused on predicting market trends. The vast array of information on social media regarding public sentiment towards current events, coupled with the known impact of financial news on stock prices, has led to the application of data mining techniques for understanding market volatility. This research proposes a novel method that integrates social media data, encompassing public sentiment, news, and historical stock prices, to predict future stock trends. The approach involves two primary phases. The first phase develops a sentiment analysis (SA) model using three dilated convolution layers for feature extraction and classification. Addressing the challenge of unbalanced classification, a reinforcement learning (RL)-based strategy is employed, wherein an agent receives varied rewards for accurate classification, with a bias towards the minority class. Additionally, a unique clustering-based mutation operator within a differential equation (DE) framework is introduced to initiate the backpropagation (BP) process. The second phase incorporates an attention-based long short-term memory (LSTM) model, merging historical stock prices with sentiment data. An experimental analysis of the study dataset is conducted to determine optimal values for significant parameters, including the reward function.**

Keywords—Social media; stock market; sentiment analysis; unbalanced classification; reinforcement learning; differential equation; long short-term memory

I. Introduction

Before the World Wide Web, people's decisions were influenced by opinions from colleagues and friends. With the rise of the internet, there's been an increase in sharing opinions with strangers online. People now express their views on a variety of topics, including news and products, on social networking sites like Twitter [1]. SA is a research field focused on categorizing these opinions into positive, negative, or neutral sentiments, often involving opinion mining to analyze textual expressions. Social media platforms like Facebook and Instagram have become key for expressing opinions, offering a vast resource for SA to gain insights into public sentiment on diverse topics [2].

In recent years, a specialized market has developed, focusing specifically on detecting and analyzing sentiments in the financial sector. This niche is centered on identifying and assessing sentiments tied to financial transactions. The skill to predict stock prices accurately is highly valuable for researchers and investors, especially given the volatility and unpredictability of these prices [3]. Machine Learning (ML) techniques have shown potential in improving the accuracy and dependability of stock market forecasts, yet designing an effective stock prediction model is still a complex task. Stock market behavior is shaped by a mix of social mood and historical price trends, which significantly influence stock price movements and changes. Integrating these elements into predictive models is essential for obtaining meaningful insights and making well-informed investment choices.

The influence of daily news articles on stock market performance is substantial [4]. These articles, containing vital information about companies, budgets, and trends, significantly affect public sentiment and stock market strategies. By analyzing these articles, investors can obtain critical insights for better investment decisions. This paper aims to use news articles to forecast stock market trends, focusing on those that provide insights into specific industries and predict future stock price movements. The increasing availability of financial data offers an opportunity to improve the speed and accuracy of these predictions [5].

SA faces challenges due to data imbalance, marked by significant discrepancies between negative and positive instances [6]. To combat this, there are both algorithm-based and data-based approaches. Data-based techniques include under-sampling, over-sampling, or combining both to reduce the impact of class imbalance. For instance, SMOTE [7] creates new samples by interpolating between minority instances, while NearMiss [8] applies under-sampling through the nearest neighbor algorithm. While over-sampling may lead to overfitting, under-sampling could result in the loss of important information. On the algorithm side, strategies are developed to give more weight to the underrepresented class. These involve enhanced ensemble learning, adjusting decision thresholds, and implementing cost-sensitive learning methods [9]. In cost-sensitive learning, classification is seen as a cost-minimization problem, assigning greater penalties for misclassifying minority samples. Ensemble techniques combine the predictions of multiple classifiers, and threshold adjustments modify the decision threshold during the testing phase.

Deep learning techniques, including Deep Reinforcement Learning (DRL), offer effective solutions for imbalanced classification issues [10, 11]. DRL stands out for its ability to manage imbalanced data through its distinctive features. It employs a reward system that prioritizes the minority class by either penalizing misclassification more severely or rewarding correct identification more generously. This method proactively counteracts the tendency of traditional models to favor the majority class. The benefits of DRL extend beyond just equalizing class distribution. It improves the recognition of significant patterns, especially those pertaining to the minority class, by adeptly filtering out irrelevant data. DRL's capacity to identify key, often-missed features in the dataset plays a vital role in developing models that are both more accurate and efficient [12].

Neural network weight initialization significantly affects the training efficiency and accuracy in SM prediction. Usually, weights are randomly assigned in gradient-based training algorithms like backpropagation [13, 14]. However, the choice of initial weights is vital for effective convergence and overall training performance [15, 16]. To improve weight initialization, population-based training can be used, selecting the optimal model from a pool to initiate the neural network. This method helps overcome the issue of local optima common in traditional methods. Notably, evolutionary algorithms have shown to be as effective as stochastic gradient descent for neural network training [17, 18]. Differential Evolution (DE) [19], a population-based optimization algorithm, is effective for machine learning weight initialization [20]. It offers several benefits: it explores the solution space with diverse candidate solutions, avoiding local optima and leading to optimal weight configurations. DE's iterative update process, based on the differential between target and current solutions, results in faster convergence and better performance. Additionally, DE is robust against noisy data, ensuring stable initial weight configurations even amidst data uncertainties. Its flexibility allows for customization to specific needs, like setting weight boundaries or incorporating prior knowledge, making it adaptable for various learning tasks and enhancing its effectiveness in weight initialization.

In this study, a novel approach is introduced to analyze social media data, combining public sentiment, personal opinions, news trends, and historical stock data to predict future stock prices. The methodology consists of two main stages. In the first stage, a SA model is developed using three dilated convolution layers to extract feature vectors and perform classification tasks. To address the challenge of unbalanced classification, a unique strategy based on RL is proposed, treating the task as a series of decision-making processes. The agent receives rewards at each step based on accurate classification, with a smaller reward assigned to the majority class to address the class imbalance. In the second stage, an attention-based Long Short-Term Memory (LSTM) approach is employed. This involves combining historical stock prices with the sentiment analysis results from the previous stage to make informed predictions. The optimal values for key parameters, including the reward function, are determined through experimental studies conducted on the dataset. To assess the individual contributions of the RL component, ablation studies are conducted to confirm the independent and cumulative positive effects on the overall performance of the model. The key contributions of the proposed model can be distilled into the following points:

- In the SA model, an ensemble of dilated convolutions is employed to extract valuable insights from textual data. This enhances the accuracy and informed decision-making in classification tasks.
- To address the challenge of imbalanced classification in SA, an RL strategy is proposed. This innovative approach provides a fresh perspective on achieving a balance between different classes.
- The model incorporates a unique reward system that reinforces correct decisions and penalizes incorrect ones. By assigning higher rewards to the minority class, the issue of dataset imbalance is effectively tackled, encouraging the model to give due attention to underrepresented data. This strategic approach contributes to a more balanced and equitable classification process.
- An improved DE algorithm that utilizes clustering to initialize weights in the SA model effectively has been devised. This approach aids in identifying a promising region to initiate the BP algorithm within the model. By selecting the optimal or near-optimal solution from the best cluster as the initial solution in the mutation operator and employing a new updating strategy, candidate solutions are generated more efficiently.

The structure of the article is as follows: Section II provides a literature review on stock market prediction. In Section III, the proposed approach is delved into in more detail. The experimental results and analyses are presented in Section IV. Finally, Section V concludes the paper.

II. RELATED WORK

Opinion mining focuses on capturing individuals' unique perspectives and opinions and finds applications in various fields such as product reviews, surveillance, healthcare, and politics. Accurately predicting changes in stock prices is a crucial research area, and recent advancements have been made in creating predictive models for the global stock market. Traditional methods like time series analysis and machine learning (ML) models are commonly used in academic research to analyze stock market predictions. Numerous approaches for analyzing social media (SM) have been explored in scholarly literature. Milosevic et al. [21] and Pandya et al. [22] introduced an ML method that utilizes selected financial variables to predict long-term investments, achieving improved performance through feature selection. Porshnev et al. [23] combined support vector machine (SVM) and neural network (NN) algorithms, utilizing a lexicon-based approach to analyze psychological states and their impact on stock market indicators like DJIA and S&P500. They effectively demonstrated the influence of Twitter data on stock market performance. Lai et al. [15] investigated stock prediction models using support vector machine (SVM) and least square SVM. Athale et al. [24] and Mehta et al. [25] utilized SA and ML to explore the relationship between public sentiment and stock performance, aiming to address the complex and unpredictable nature of non-linear and non-parametric financial time series. Xing et al. [26] highlighted the capabilities of natural language processing (NLP) in financial forecasting, particularly in the domain of Natural Language-Based Financial Forecasting (NLFF) or sentiment analysis using data from SM platforms. NLP techniques are rapidly advancing, specifically in NLFF and the analysis of SM data, driven by practical applications in financial forecasting. However, despite these advancements, a gap remains in integrating comprehensive social media data with advanced sentiment analysis and machine learning techniques for precise stock market predictions. The proposed research aims to bridge this gap by combining public sentiment, news, and historical stock prices to predict future stock trends more accurately.

Emotion plays a crucial role in facilitating effective communication, as indicated by research conducted by Pandya et al. [22] and Smailović et al. [27], who explored the usefulness of Twitter feeds for predicting stock closing prices and assessing public sentiment towards companies and their products. In this paper, a method is introduced to measure the probability of positive and negative expressions or opinions, with the aim of forecasting SA in the field of finance. Using the Granger causality test, stock price trends can be examined over a short period, providing valuable insights from the data [5]. The use of SVM enables the categorization of tweets into positive, negative, or neutral sentiments, thereby enhancing the predictive capability of SM platforms. Various initiatives have been undertaken to predict SM trends, with a focus on accurately forecasting the significant impact of a company's SM presence through real-time collection of Twitter data. The study demonstrates the effectiveness of sentiment analysis in extracting public mood from Twitter and other SM platforms to predict fluctuations in individual stock prices. In order to bolster the analytical process, the active learning model has been integrated with a stream-based method. This combination enables the algorithm to choose fresh training data, thereby enhancing its performance. The financial implications of this analysis are evaluated by employing Recurrent Neural Networks (RNNs) in a trial-based methodology. The proposed model extends these efforts by employing a novel clustering-based mutation operator within a DE framework and integrating sentiment analysis with attention-based LSTM models, a distinct approach not yet explored in existing literature.

Examining sentiment helps in evaluating the impact of emotions within a textual context when analyzing decision-making with a positive outlook. Bhuriya et al. [28] developed predictive models using regression techniques to forecast the market price of TCS. These models were based on attributes such as large price, close price, open price, small price, and volume. The study assessed the effectiveness of linear, polynomial, and radial base function regression models by considering the confidence values associated with the predicted results. Despite having expertise, predicting stock prices remains challenging due to unexpected fluctuations influenced by historical trends and past price changes in the global market economy. Barot et al. [4] proposed a research approach that utilized sentiment analysis to analyze news articles and forecast stock price fluctuations. They introduced a method for categorizing articles based on sentiment, classifying them as positive, negative, or neutral. This approach was integrated into machine learning models to improve the accuracy of stock market prediction. Patel et al. [29] and Patel et al. [30] examined the practical implementation of a machine learning model to forecast stock and share price movements in the Indian equity market. The analysis incorporated four forecasting models (SVM, ANN, Naïve Bayes, and Random Forest) with two input methods. Khedr et al. [31] introduced an optimized model that aimed to reduce error ratio and improve the accuracy of predicting patterns in share price performance. Their predictive model integrated sentiment analysis of financial news and historical values of social media data. These approaches have successfully utilized various types of market and company data, leading to superior outcomes compared to previous research. Chen et al. [32] conducted a comparison between conventional neural network price prediction and deep learning techniques using Chinese social media data on CSI 300 shared values. Their findings indicated that deep learning predictions outperformed conventional neural networks in terms of performance. Carosia et al. [33] conducted a study to investigate the influence of SM activities, specifically on the Brazilian SM platform Twitter, on the market value of specific companies. They utilized sentiment analysis (SA) to analyze the impact of SM movements on these companies. The study considered three different perspectives: the total count of emotions expressed in tweets, tweet sentiments weighted by the number of favorites, and tweet sentiments weighted by the number of retweets. SA was performed using the Multilayer Perceptron technique to analyze sentiment in Portuguese. Deep learning algorithms, including deeply convolutional neural networks (CNN), have gained prominence in SA. Santos et al. [34] employed a deep CNN trained on the Stanford Sentiment Treebank and the Stanford Twitter Sentiment Corpus. The model achieved an impressive accuracy of over 85% in predicting sentiment for both datasets. Yoshihara et al. [35] discussed an SA technique that combined recurrent neural networks (RNN) with Deep Belief Networks (DBN) in deep learning, resulting in improved financial market forecasting with reduced error rates compared to SVM and DBN strategies. In another study by Jiang et al. [36], deep learning models were analyzed to forecast SM trends. The research involved categorizing various NN models, evaluating metrics, exploring their integration, and assessing their reliability. Pang et al. [37] and Sun et al. [38] proposed the utilization of STM in NN techniques to enhance the accuracy of financial market predictions, particularly for the Shanghai A-shares composite index. Khan et al. [39] incorporated data algorithms that combined social networks and business news to evaluate the impact on the accuracy of SM forecasts over a 10-day period. They implemented feature selection and spam tweet removal techniques, and the use of deep learning techniques improved the accuracy rates. The research findings indicated that assessing the impact of SM is a complex task, with New York and Red Hat stocks showing significant sensitivity to SM activity, while London and Microsoft stocks are more influenced by financial news.

To sum up, existing literature showcases a variety of methods for price prediction. Although traditional deep learning techniques have significantly progressed the field, their limitations include a lack of social media data utilization and sensitivity to initial weights, which hinders their practical application in forecasting. Moreover, the issue of unbalanced classification remains a prevalent challenge for many deep learning models. Addressing these shortcomings, the study introduces an innovative methodology that combines the strengths of RL with the adaptive potential of a DE algorithm, specifically fine-tuned for initial weight optimization. This approach is engineered to surmount the conventional obstacles encountered by deep learning methods, thereby boosting the model's adaptability and efficiency in identifying cost-effective strategies for price prediction. The goal is to offer a sophisticated and robust tool to the suite of price prediction methods, one that is not only theoretically ground-breaking but also practically relevant in a variety of real-world scenarios.

III. Proposed Method

Based on Fig. 1, the proposed model is structured into four distinct steps, each integral to the overall process:

- The first step is Data Collection, where relevant data is gathered from various sources.
- The second step is Data Pre-processing. In this phase, the collected data undergoes cleaning and transformation to make it suitable for analysis.
- In the third stage, the focus shifts to analyzing news content using NLP for SA. This essential phase leverages advanced natural language processing (NLP) methods to assess the emotional undertones within the news material. The aim is to systematically identify the sentiment of each article, classifying it as either positive, negative, or neutral. The SA model is specially tailored to refine the classification mechanism, addressing critical challenges such as class imbalance and the need for accurate initial weight configurations. The integration of DE and RL in the proposed framework specifically targets these pivotal areas, which are often inadequately addressed by existing models. Conventional techniques typically lack a structured approach for setting initial weights, potentially slowing down the learning process and leading to suboptimal solutions. In the context of SA, where promptness and precision in prediction are essential, these limitations can be significant. Moreover, the model's RL component is designed to provide greater incentives for correctly predicting less frequent classes, thus shifting the focus to these vital predictions. This represents a marked advancement over traditional supervised learning models, which may struggle with insufficient data representation across all categories. The flexibility of RL in adjusting the learning strategy ensures a more equitable exploration of decision-making options, fostering approaches that prioritize the accurate identification of less represented categories. This distinctive capability of RL within the model distinguishes it from current methods, enabling it to address the unique challenges that traditional classification models face in SA applications.
- The final step is SM forecasting. Using the insights gained from the sentiment analysis, along with other relevant financial indicators and historical data, this stage focuses on predicting future movements of the stock market.

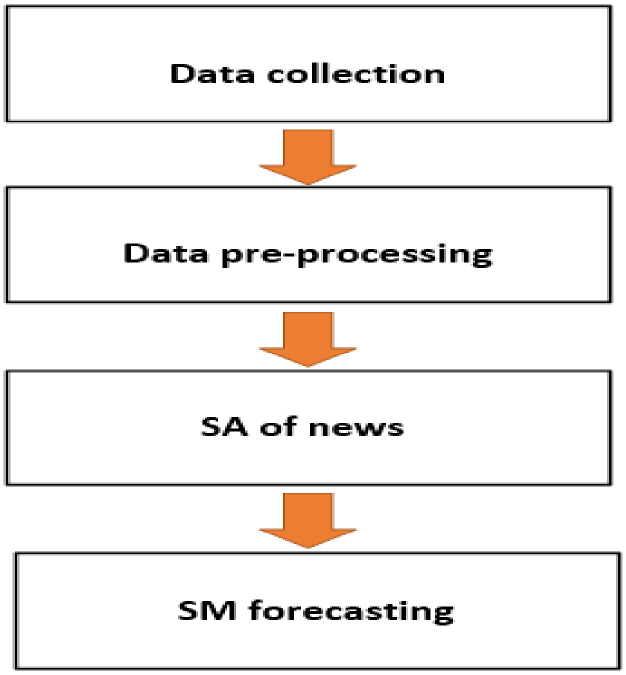


Fig. 1. The constituent stages of the proposed model.

A. Data Collection

The system design is founded on the use of two distinct datasets from various data providers. The first dataset includes news articles from reputable platforms like Moneycontrol, IIFL, Economic Times, and Twitter. This collection provides a broad perspective on market-related news, encompassing company announcements, industry trends, and economic developments, crucial for understanding stock market behavior. The second dataset contains historical records from the National Stock Exchange (NSE) of India, spanning six years. It features key parameters such as date, time, open, high, low, close values of stocks, and trading volume. This dataset is instrumental in analyzing past market trends and patterns, thereby aiding in generating insights into the behavior of individual stocks and the overall market. The data management strategy involves storing the gathered information in CSV file format. This approach offers several benefits. Firstly, it ensures convenient organization, enabling users to navigate and interpret the data efficiently. Secondly, the reduced file size makes it more manageable and less resource-intensive to handle large amounts of data. Additionally, the ease of generating and manipulating CSV files facilitates smooth data processing. Finally, this format enhances data integrity, ensuring that the information remains accurate and reliable for analysis. By integrating these datasets, the system aims to create a comprehensive framework that leverages the strengths of both news and stock market data. This dual-dataset approach is expected to provide a more nuanced and detailed understanding of the factors influencing the stock market, thus supporting more informed and effective decision-making in financial analytics.

B. Data Pre-processing

The pre-processing phase of the study begins with synchronizing text data and converting it to lowercase characters. This step is vital to address various text factors that could impact the classification process. The primary objective is to prepare the input data for sentiment classification, making it suitable for analysis. This preparation involves several tasks, such as removing links, special symbols, and emoticons, along with eliminating stop words. Further, the process includes analyzing parts of speech, applying stemming techniques, and tokenizing the text. These steps are essential to extract meaningful features for sentiment analysis. The output from the pre-processor subsystem then becomes the input for the sentiment classifier method.

The framework described in the article adopts a data-driven approach, leveraging historical stock prices recorded at opening and closing times. This rich dataset forms the foundation for analyzing and understanding stock price behavior over time. To represent these prices effectively, the framework utilizes a sentiment polarity approach. This involves categorizing stock prices as either positive or negative, aiming to capture the underlying emotional sentiment in the stock market. The sentiment polarity representation is instrumental in providing insights into market dynamics, highlighting the influence of emotions and perceptions on stock prices. To substantiate the effectiveness of this approach, a study was conducted at Stanford University [40]. This study integrated sentiment polarity values derived from the framework, aiming to explore the correlation between sentiment polarity and stock price movements. The findings of this study are enlightening, revealing the intricate relationship between market sentiment and financial outcomes. They underscore the importance of sentiment analysis in predicting stock market trends. These insights are particularly beneficial for investors, traders, and financial analysts, aiding them in making more informed and strategic decisions.

C. SA of News

Fig. 2 in the study illustrates the proposed structure for the SA model. This model functions based on a sentence D, defined as a series of words $[w_1, w_2, \ldots, w_\mathrm{n}]$, where n is the maximum number of words in a sentence. The sentence is fed into the BERT model, which generates an embedding matrix $E = [e_1; e_2; \ldots; e_n]$, with each e_i representing the embedding of the corresponding word w_i. To extract features, three dilated convolution layers are applied in parallel to matrix E. Each of these layers independently extracts a distinct feature vector from the sentence. Following this, max pooling is implemented to capture the most significant features while simultaneously reducing computational complexity. The outputs from the max-pooling layers are then directed into a Multilayer Perceptron (MLP) for classification purposes. The MLP's output is a vector of length three, representing the classification of the input sentence into one of three categories: positive, negative, or neutral. However, one of the challenges encountered in this model is the imbalanced nature of the dataset, predominantly skewed towards sentences classified as positive. This imbalance can potentially impair the performance of the system. To address this imbalance, the model incorporates an imbalanced classification Markov decision process. This process involves the use of a sequential decision-maker, specifically designed to tackle the challenges posed by the imbalanced classification problem. The decision-maker operates by adjusting the classification strategy, focusing on equitably distributing attention across all classes, including those less represented in the dataset. This approach not only enhances the accuracy of the model but also ensures a more balanced and equitable classification outcome, vital for the integrity and applicability of sentiment analysis in various contexts.

1) Pre-training: Weight initialization plays a vital role in the performance of deep learning models [41]. Choosing the right initial values for weights is crucial because inappropriate values can lead to convergence problems, making the model less effective or even rendering it non-functional. Proper initialization helps in achieving a faster convergence rate and improves the overall training efficiency of the model. It also impacts the ability of the model to reach global or good local minima in the optimization landscape. Different initialization methods, such as Xavier/Glorot, He initialization, or random initialization, are tailored for specific types of neural networks and activation functions. These methods aim to maintain a balance in the variance of the activations and gradients throughout the network, which is critical for deep models. The right choice of weight initialization method can significantly influence the learning dynamics and the eventual success of the model in tasks like classification, regression, or feature learning.

DE [42] stands as a prominent method within the realm of evolutionary algorithms, primarily used for solving complex optimization problems. As a population-based approach, DE focuses on iteratively refining a group of candidate solutions to converge towards an optimal solution for a given problem. The strength of DE lies in its unique mechanisms of mutation, crossover, and selection, which collectively drive the evolutionary process. Mutation in DE is a distinctive process where a new candidate solution is generated by adding the weighted difference between two randomly selected solutions from the population to a third solution. This approach introduces diversity and aids in exploring the solution space. The crossover step in DE further enhances solution diversity by combining features of the mutated solution with an existing member of the population, creating a trial solution [43]. This crossover mechanism ensures a mix of characteristics from different solutions, promoting the exploration and exploitation of the solution space. Selection, the final step in DE, is critical for the evolution of the population. It involves a comparison between the trial solution and the existing solution, retaining the one that offers better performance according to the defined objective function. This selective pressure gradually improves the overall quality of the population, steering it towards the optimal solution. DE's effectiveness is not limited to a single domain but extends across various fields such as engineering, economics, and machine learning. Its popularity stems from its simplicity, efficiency, and versatility in handling different types of optimization problems, including those with nonlinear, multimodal, and high-dimensional characteristics. Its ability to find high-quality solutions with relatively simple operations and fewer control parameters makes it a go-to choose for practitioners and researchers in optimization tasks.

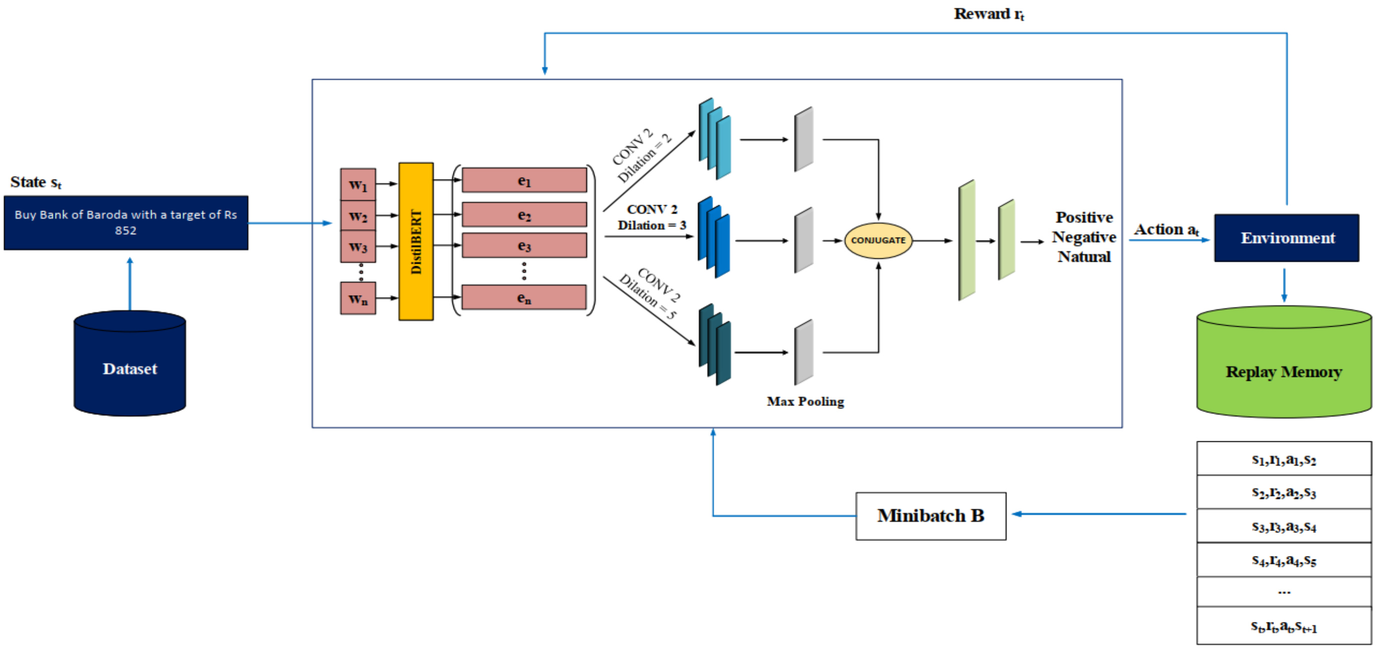


Fig. 2. The proposed SA model.

The mutation operator in DE is a pivotal component that significantly influences the algorithm's ability to search effectively across the solution space. This operator is responsible for introducing variability in the population, which is essential for exploring new regions in the search space and avoiding local optima. However, the effectiveness of the mutation operator is a delicate balance. If it is overly aggressive, it might disrupt the population's structure too much, leading to the loss of promising solutions and potential premature convergence or stagnation. In contrast, an overly conservative mutation approach may not provide sufficient exploration, leading to slow convergence rates and possibly settling for suboptimal solutions. To tackle these challenges, various modifications and enhancements to the standard mutation process in DE have been proposed. One of the notable advancements is the introduction of adaptive mutation schemes. These schemes dynamically adjust the mutation parameters, like the mutation rate and scaling factor, based on real-time feedback from the population's performance. By doing so, the algorithm can maintain an optimal balance between exploration and exploitation throughout the optimization process, adapting to the changing landscape of the search space. Another innovative approach is self-adaptive DE, where the mutation parameters are not fixed but are treated as part of the solution itself. In this method, each candidate solution carries its mutation rate and scaling factor, which evolve alongside the solution. This self-adaptation allows for a more tailored mutation process for each candidate solution, potentially enhancing the diversity and adaptability of the population [44].

To further boost the performance of the DE algorithm, a cutting-edge mutation and updating approach, which incorporates clustering concepts, is utilized to bolster the optimization process. This innovative strategy, drawing inspiration from the methodologies outlined in reference [45], focuses on employing the mutation operator to precisely target designated sectors within the search landscape. This is achieved through the application of the k-means algorithm, which effectively divides the current population P into k distinct clusters. Each cluster corresponds to a unique segment of the search space, thus facilitating a more focused and efficient exploration. The determination of the number of clusters is conducted in a stochastic manner, with the possible range stretching from 2 to the square root of N ($\sqrt{N}$). Upon completion of the clustering phase, the algorithm proceeds to identify the most optimal cluster. This is determined based on the average fitness level of the members within each cluster, with the one displaying the lowest mean fitness being designated as the most favorable. This particular cluster is then prioritized in the subsequent phases of the algorithm, as it is indicative of a region in the search space with a higher potential for yielding optimal solutions. Fig. 3 serves as a visual representation of this advanced methodology. In this figure, a theoretical problem encompassing 19 potential solutions is illustrated. These solutions are methodically grouped into three separate clusters, each representing a different area of the solution space. This visualization not only demonstrates the clustering mechanism but also highlights the effectiveness of this approach in partitioning the population for more targeted and efficient optimization. Through this sophisticated clustering-based mutation and update strategy, the DE algorithm is significantly enhanced, offering a more robust and precise tool for tackling complex optimization challenges.

The proposed mutation, based on clustering, is then defined as [45]:

$$\overrightarrow{v_i^{clu}} = \overrightarrow{win_g} + F(\overrightarrow{x_{r_1}} - \overrightarrow{x_{r_2}}) \quad (1)$$

In this context, $\overrightarrow{x_{r_1}}$ and $\overrightarrow{x_{r_2}}$ symbolize two randomly chosen candidate solutions from the existing population, whereas $\overrightarrow{win_g}$ represents the optimal solution found within the identified promising region. It is critical to acknowledge that $\overrightarrow{win_g}$ may not always equate to the best solution across the entire population.

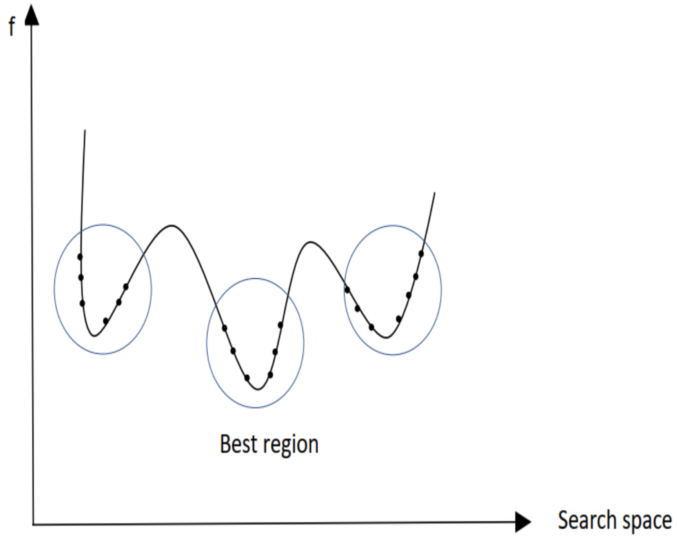


Fig. 3. Utilizing population clustering within the search space to identify the most favorable region.

Subsequent to the creation of M novel solutions via the mutation process influenced by clustering, the current population undergoes an update process as per the guidelines of Gradient-based Population Adjustment (GPBA) [46]. GPBA is a sophisticated mechanism designed to refine the population in evolutionary algorithms like DE. GPBA works by evaluating the gradient information of the current population. This involves assessing how each candidate solution is positioned relative to the objective function's gradient. By doing so, GPBA can determine the direction in which the solutions should be adjusted to move closer to the optimal point in the search space. This gradient-based approach differs from traditional evolutionary strategies, which rely solely on fitness-based selection and random mutations. The key advantage of integrating GPBA into the DE algorithm is the enhanced convergence speed towards optimal solutions. Since GPBA utilizes gradient information, it can guide the population more effectively towards the global optimum, especially in complex, multimodal landscapes where traditional methods might struggle. In the context of the DE algorithm, after the clustering-based mutation process identifies promising regions and generates new solutions, GPBA takes over to fine-tune these solutions. It adjusts the population by moving the candidate solutions along the gradient of the objective function. This not only accelerates the convergence process but also helps in maintaining diversity within the population, preventing premature convergence to local optima.

The procedure for this update is outlined as follows:

- Selection: Generate k random individuals to serve as the initial seeds for the algorithm.
- Generation: Produce a set of M solutions using mutation based on clustering and denote it as v^{clu}.
- Replacement: Choose M solutions randomly from the current population to form set B.
- Update: Choose the best M solutions from the union of sets v^{clu} and B to form set B'. The new population is obtained by combining the elements of set P that are not in B with the elements of a set B' $((P - B) \cup B')$

2) Deep q-network training: In sentiment analysis, unbalanced classification occurs when the distribution of sentiment labels in the dataset is highly skewed, with one sentiment class being more prevalent than the others. This poses challenges as standard classification algorithms tend to favor the majority class, leading to poorer performance for the minority class [47, 48]. Consequently, models might excel at predicting the majority sentiment but struggle with accurately identifying and classifying minority sentiments. The imbalanced nature of the dataset can introduce bias and impact the learning process, resulting in models with lower recall or sensitivity for the minority class. This is particularly problematic when the minority class represents important sentiments, such as negative feedback in customer analysis. To address the challenge of imbalanced classification, a sequential decision-making strategy utilizing RL is adopted. In RL, an agent engages with its environment, aiming to maximize cumulative rewards through the selection of optimal actions. Within the devised model, the agent acquires a sample from the dataset and undertakes a classification task at each sequential time step. Subsequently, the agent obtains immediate feedback from the environment: accurate classifications are rewarded with positive scores, while erroneous classifications incur negative scores. This approach ensures dynamic learning and adaptation by the agent, enhancing its decision-making capabilities in imbalanced classification scenarios [49, 50]. Suppose a dataset of N samples is available, each with corresponding labels $D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_N, y_N)\}$, where x_i represents a sample and y_i denotes its label. The upcoming details provide the planned arrangements for the proposed approach.

- Policy π_θ : Policy π serves as a function which links states (S) to corresponding actions (A), with $\pi_\theta\ (s_t)$ denoting the action taken in a particular state s_t. The classification technique that employs the weights θ is identified as π_θ.

- State s_t : Each sample x_t from the dataset, D is associated with a specific state s_t. The beginning state s_1 is symbolized by the first piece of data x_1. To deter the model from acquiring a particular sequence, D is shuffled in every episode.
- Action a_t: The action a_t is performed to estimate the label x_t, where the categorization is binary, and a_t can take on either the value of 0 or 1. In this context, 0 signifies the less prevalent class, whereas 1 symbolizes the more common class.
- Reward r_t : The reward is determined by how well the action is executed. If the agent correctly classifies, it is given a positive reward; conversely, if it errs, it is penalized with a negative reward. The reward value should be varied for each class. By appropriately calibrating rewards, the model's performance can be significantly improved by ensuring that the reward magnitude matches the corresponding action. The method for determining the reward for a given action in this study is described by the following equation:

$$r_t(s_t, a_t, y_t) = \begin{cases} +1\, , a_t = y_t \ and \ s_t \in D_{Maj} \\ -1\, , a_t \neq y_t \ and \ s_t \in D_{Maj} \\ \lambda\, , a_t = y_t \ and \ s_t \in D_{Min} \\ -\lambda\, , a_t \neq y_t \ and \ s_t \in D_{Min} \end{cases} \tag{2}$$

Here, D_{Maj} and D_{Min} denote the majority and minority classes correspondingly. Accurately or inaccurately classifying a sample from the majority class results in a reward of $+\lambda$ or $-\lambda$, respectively, where λ is a value between 0 and 1.

- Terminal E: Throughout every training episode, the training process culminates at different terminal states. A series of state-action pairs $\{(s_1, a_1, y_1), (s_2, a_2, y_2), (s_3, a_3, y_3), \dots, (s_t, a_t, y_t)\}$ from the starting state to the ending state is known as an episode. In the context of the situation, the conclusion of an episode is marked either by classifying all the training data or by misclassifying a sample from the minority class.
- Transition probability P: The agent advances from the present state, s_t, to the subsequent state, s_{t+1}, following the order of the data being read. The likelihood of this transition is denoted as $p(s_{t+1}|s_t, a_t)$.

D. SM Forecasting

LSTM networks have garnered considerable attention in stock market analysis due to their ability to capture and leverage temporal relationships in time series data [51]. Unlike traditional statistical models, LSTM networks excel at learning and adapting to complex patterns and dependencies in the data, making them well-suited for predicting stock market prices or trends. One advantage of LSTM networks is their proficiency in handling long-term dependencies, enabling them to capture subtle relationships and trends that span extended periods. By analyzing historical price data, trading volumes, and other relevant factors, LSTM networks can identify recurring patterns, seasonal trends, and market cycles that traditional models might miss [52]. Another strength of LSTM networks is their capacity to process and retain information over long sequences. The architecture of an LSTM cell includes memory units that selectively remember or forget information over time, enabling the network to store and retrieve relevant historical data while disregarding irrelevant or redundant information. This memory retention mechanism is crucial in capturing the dynamics of stock market movements, where past trends and events significantly impact future outcomes.

Additionally, LSTM networks can effectively handle irregularities and non-linearities in stock market data. The stock market is influenced by various factors, such as economic indicators, geopolitical events, news, and investor sentiment, leading to abrupt shifts and anomalies in stock prices. LSTM networks, with their ability to model complex non-linear relationships, can effectively capture and respond to these non-linear dynamics, aiding in market prediction.

The LSTM model, initially conceptualized by Hochreiter and Schmidhuber [53], has evolved significantly through various advancements since its inception [54]. In this particular study, the focus is on the highly regarded LSTM architecture developed by Gers et al. [55], a structure that has seen extensive application in diverse academic research [56, 57]. The fundamental aspect of the LSTM model is its distinctive gating mechanism, which effectively manages the storage and retrieval of information over temporal sequences. This research specifically explores the intricacies of the LSTM cell, in line with the delineation presented in Graves' work [36]. LSTM's architecture is characterized by a set of three specialized gates that orchestrate the efficient processing of information. These gates, denoted as i_t (input gate), f_t (forget gate), and o_t (output gate), function at a given time t in the sequence. They work in concert with c_t (memory cell) and h_t (hidden state), which are crucial components of the LSTM's internal structure. Furthermore, x_t symbolizes the external input received by the LSTM cell at the specific time t. These components collectively enhance the LSTM's capacity for learning, retaining, and applying long-term data dependencies, distinguishing it from conventional recurrent neural network designs. Its proficiency in grasping long-term dependencies renders the LSTM highly appropriate for intricate tasks associated with sequential data. This includes areas such as time series forecasting, natural language processing, and speech recognition, where context and historical data are crucial for precise modeling and prediction. The operational framework of the LSTM cell is structured in the following manner [56]:

$$i_t = \sigma(W_{xi}x_t + \ W_{hi}h_{t-1} + W_{ci}c_{t-1} + \ b_i) \tag{3}$$

$$f_t = \sigma(W_{xf}x_t + \ W_{hf}h_{t-1} + W_{cf}c_{t-1} + \ b_f) \tag{4}$$

$$c_t = f_t c_{t-1} + i_t tanh(W_{xc}x_t + \ W_{hc}h_{t-1} + \ b_c) \tag{5}$$

$$o_t = \sigma(W_{xo}x_t + \ W_{ho}h_{t-1} + W_{co}c_t + \ b_o) \tag{6}$$

$$h_t = o_t tanh(c_t) \tag{7}$$

The sigmoidal activation function, represented by the symbol σ(.), plays a pivotal role in the computational models presented. This function, which includes both logistic sigmoid and hyperbolic tangent types, is applied in a manner that targets individual elements. The weight matrices, identified as W_{xk}, where k is a member of the set $\{i, f, o, c\}$, are integral to the process, linking the input x_t with distinct components such as the input gate, forget gate, output gate, and memory cells. In addition to these, the W_{ck} weight matrices, where k falls within the set $\{i, f, o\}$, are structured as diagonal matrices. These matrices are essential in forming the connections between the memory cell and its various gates. It is vital to highlight that the neuron count designated for each gate is set beforehand. This predefined count is significant because Eq. (3) to Eq. (7), which are central to the model, are applied distinctly to each neuron, ensuring precise and individualized processing within the neural network system.

IV. Empirical Evaluation

A. Metrics

In this article, three standard metrics are used for SM prediction: root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) [58]. Each metric is defined in the following manner:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{8}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N} | \frac{y_i - \hat{y}_i}{y_i} | \tag{9}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{N} \tag{10}$$

Here, N stands for the total number of observations, with y_i representing the actual value and $\hat{y}_i$ the predicted value for each observation. Furthermore, for SA, Measures like Accuracy, F-measure, and G-means are employed, each defined in a specific manner [59]:

$$\text{Accuracy} = \frac{TP+TN}{Total\ number\ of\ samples} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{13}$$

$$Specificity = \frac{TN}{TN+FP} \tag{14}$$

$$\text{G} - \text{means} = \sqrt{\text{Recall} \times\ Specificity} \tag{15}$$

Here, TP stands for true positives, which are the actual positive cases that the model has correctly identified. TN is for true negatives, representing the actual negative instances that the model has accurately predicted. FP refers to false positives, where the model mistakenly labels actual negatives as positive. Finally, FN denotes false negatives, pertaining to the actual positive cases that the model incorrectly categorizes as negative.

B. Comparator Models

The introduced algorithm was subjected to a thorough comparative analysis against eight state-of-the-art models, including those proposed by Zhang et al. [60], Rasheed et al. [61], Jin et al. [62], Vijh et al. [63], Mehta et al. [64], Riady [65], Wang et al. [66], and BL et al. [67]. This thorough analysis was designed to provide a comprehensive view of the effectiveness of the proposed model compared to current methods. Additionally, the algorithm is compared with a derivative model called "Proposed without RL," which excludes RL for SA. The objective of this study was to validate the superior performance and effectiveness of the proposed model in the domain of stock market prediction.

C. Results

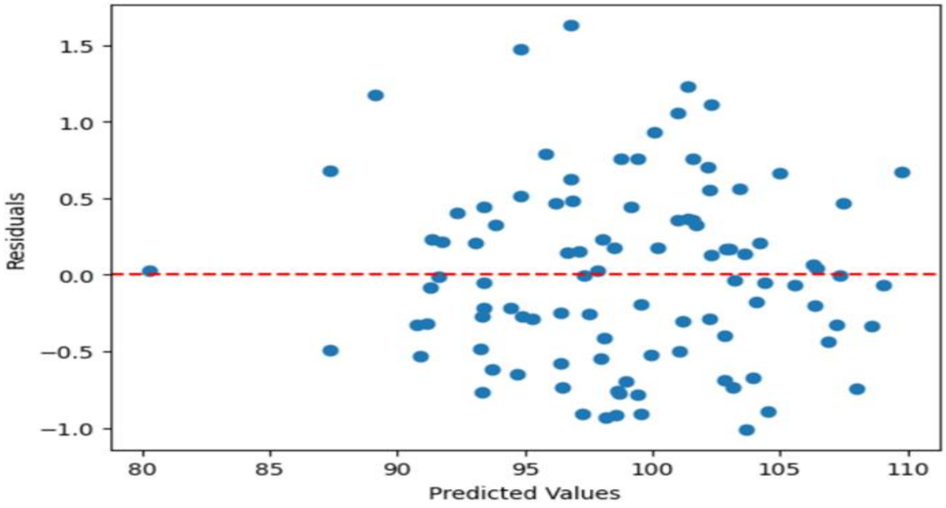
The comparative results, summarized in Table I, clearly demonstrate the exceptional performance of the proposed algorithm, consistently surpassing the other models evaluated on a shared dataset. Remarkably, the proposed algorithm outperformed the Transformer-based model by Wang et al., reducing overall error by approximately 12% and confirming the superior efficacy of attention-based LSTM over transformer. Furthermore, the model surpassed the model by BL et al., the strongest competitor, across all evaluation criteria. The performance improvement of the model is particularly significant when considering the substantial reduction in error rate. Specifically, the proposed algorithm achieved a remarkable decrease in error for two primary evaluation metrics, RMSE and MAPE, with errors reduced by over 14% and 10% respectively. This substantial reduction in error highlights the superior predictive capabilities of the algorithm. A separate comparison with the "Proposed with RL" model emphasized the critical role of RL in the proposed model. The comparison revealed an approximate 10% reduction in error rate when RL was implemented. These findings underscore the potency of RL and LSTM in navigating the complexities of time-series data, such as stock market prices, and their essential contribution to enhancing the accuracy of the proposed model.

Residual plot is a valuable graphical tool frequently employed in statistical and regression analyses. This type of plot visualizes the differences between observed and predicted values, known as residuals, on the vertical axis, while the predicted values are displayed on the horizontal axis. In Fig. 4, residual plots are presented for the models detailed in Table I.

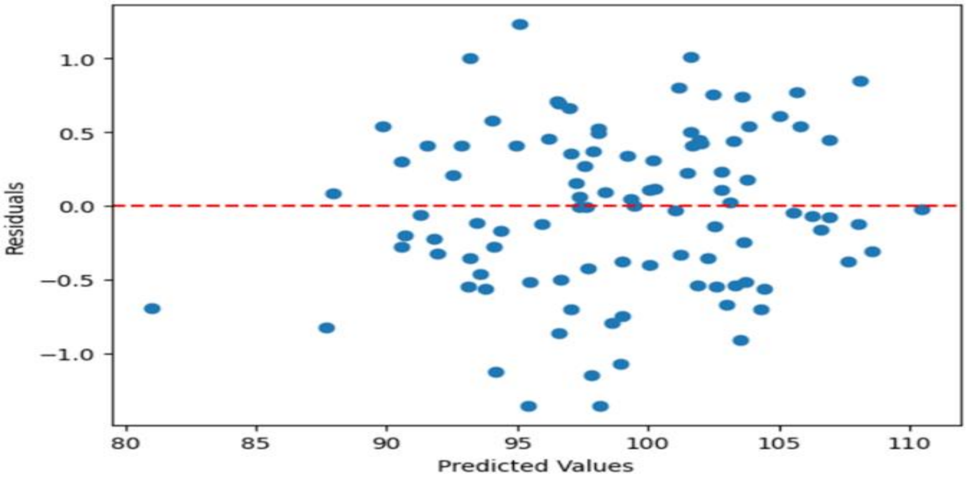
In the case of the proposed method, the data points predominantly cluster around the zero point, indicating minor discrepancies between the observed and predicted values. This clustering suggests a high level of accuracy in the model predictions. Additionally, the randomness and uniform distribution of the points demonstrate that the residuals are independent and identically distributed, a characteristic referred to as homoscedasticity. This important feature confirms that the linear regression model adheres to its assumptions, ensuring unbiased and reliable predictions. The absence of discernable patterns in the plot, such as curvilinear trends or funnel shapes, indicates that the model effectively captures the linear relationship between the predictors and the outcome variable. Moreover, this absence implies a consistent variance in the error term across various predicted value levels. These characteristics signal that the proposed model has successfully captured the underlying trends in the data and provides robust predictions.

TABLE I. RESULTS OBTAINED USING THE PROPOSED MODEL AND OTHER STATE-OF-THE-ART MODELS

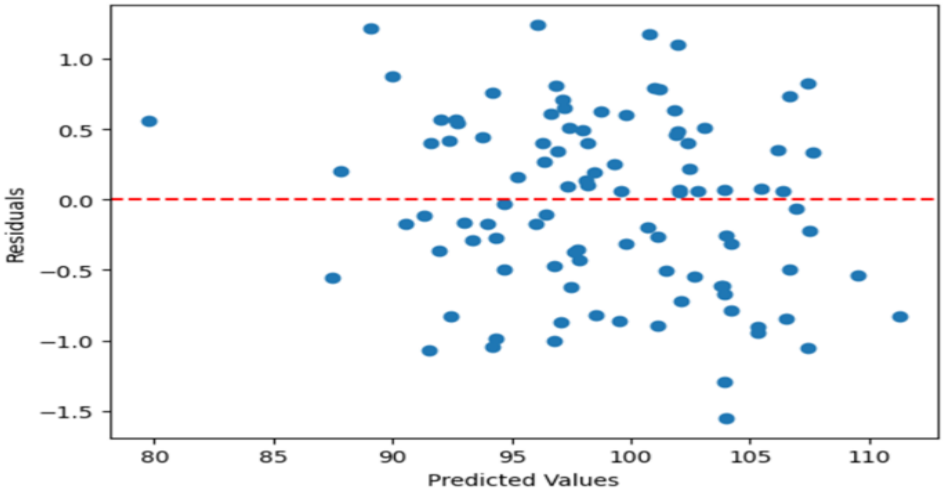
Model	RMSE	MAPE	MAE
Zhang et al. [60]	6.100	0.0410	4.100
Rasheed et al. [61]	5.402	0.0375	4.140
Jin et al. [62]	5.416	0.0302	3.715
Vijh et al. [63]	5.025	0.0260	3.493
Mehta et al. [64]	4.125	0.0241	3.201
Riady [65]	4.001	0.0217	4.128
Wang et al. [66]	3.852	0.0187	3.040
BL et al. [67]	3.501	0.0162	2.963
Proposed without RL	3.742	0.0137	2.825
Proposed	2.147	0.0125	2.130



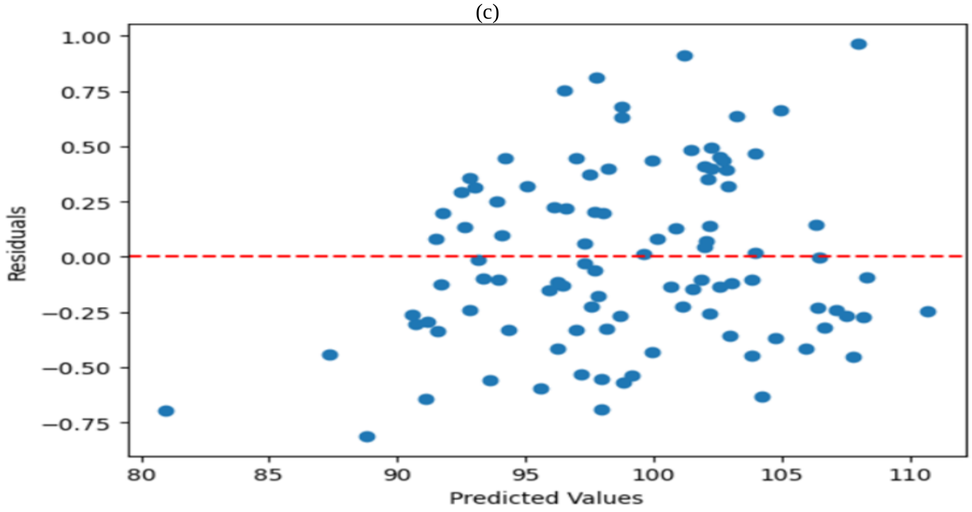
(a)



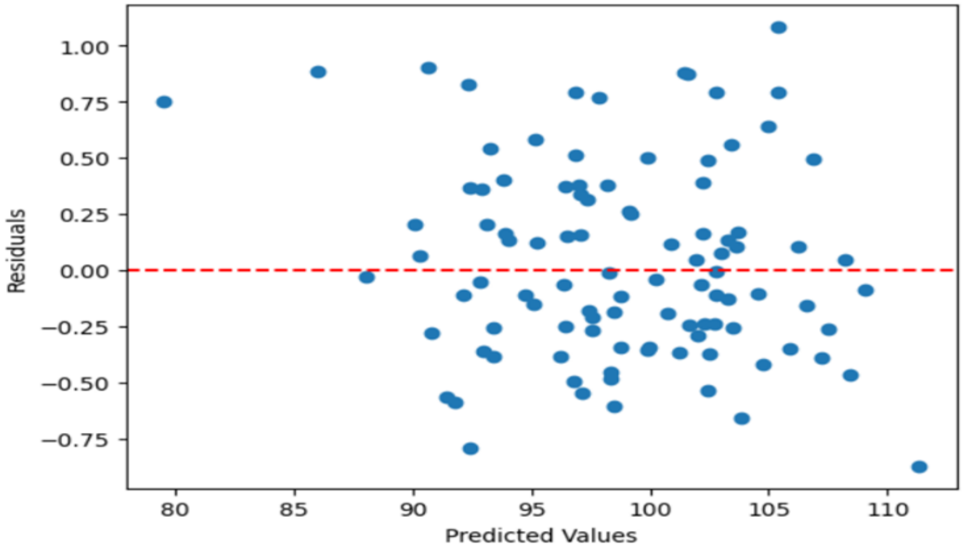
(b)


Residuals
1.0
0.5
0.0
−0.5
−1.0
−1.5
80
85
90
95
100
105
110
Predicted Values

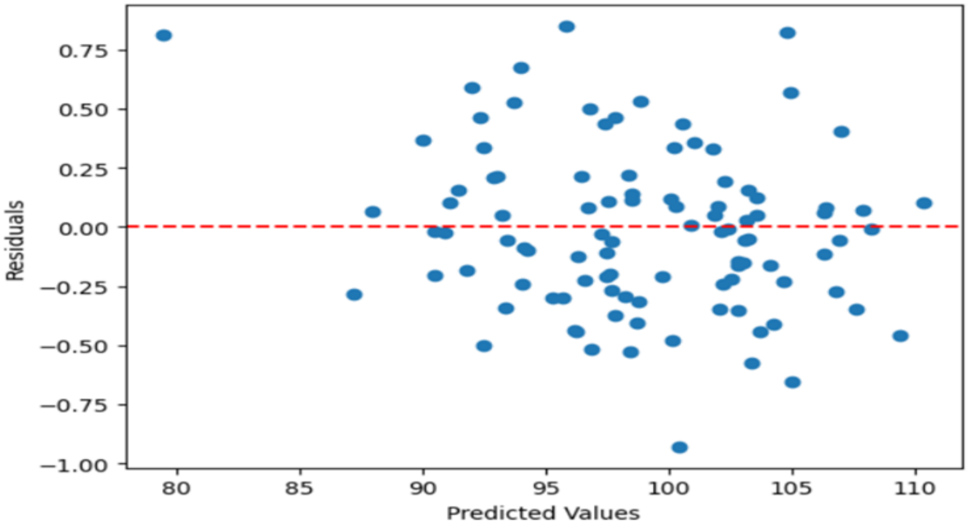
(c)


Residuals
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
80
85
90
95
100
105
110
Predicted Values

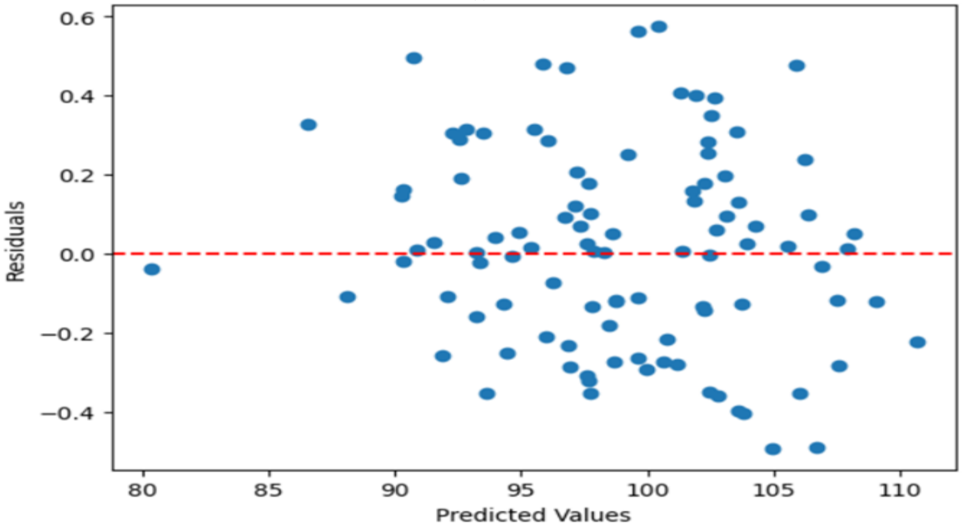
(d)


Residuals
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
80
85
90
95
100
105
110
Predicted Values

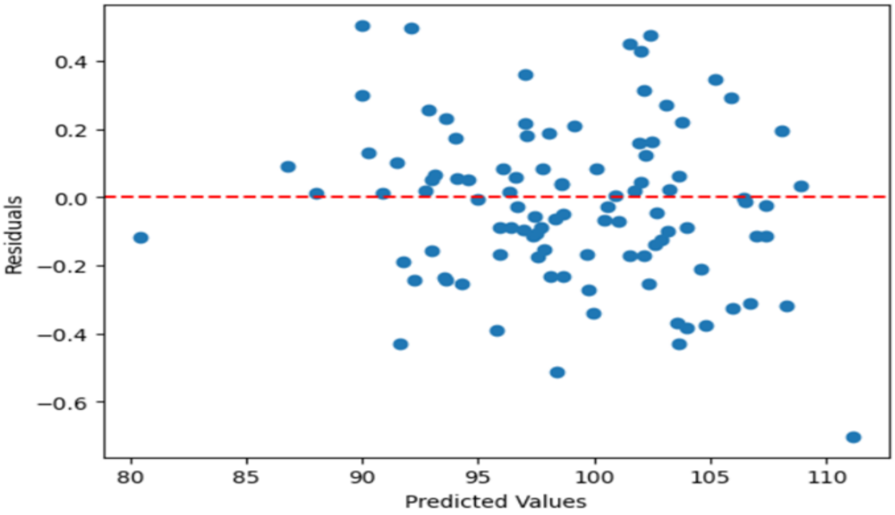
(e)


Residuals
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
80
85
90
95
100
105
110
Predicted Values

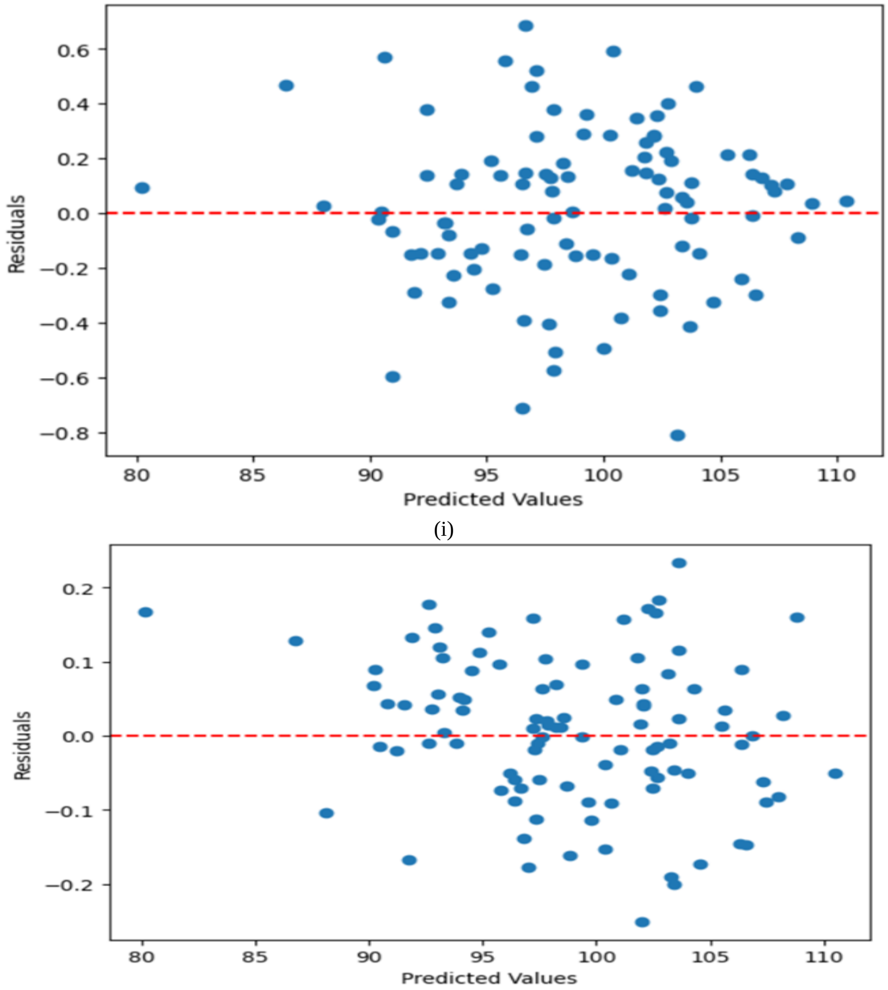
(f)


Residuals
0.6
0.4
0.2
0.0
−0.2
−0.4
80
85
90
95
100
105
110
Predicted Values

(g)


Residuals
0.4
0.2
0.0
−0.2
−0.4
−0.6
80
85
90
95
100
105
110
Predicted Values

(h)



(i)

(j)

Fig. 4. Residual plot for the proposed model and other state-of-the-art models. a) Zhang et al. [60], b) Rasheed et al. [61], c) Jin et al. [62], d) Vijh et al. [63], e) Mehta et al. [64], f) Riady [65], g) Wang et al. [66], h) BL et al. [67], i) Proposed without RL, j) Proposed.

1) Performance of semantic analysis: In the following analysis, the objective is to juxtapose the performance of the proposed SA model against five other prominent SA models. The comparative study includes models developed by Akhtar et al. [68], Akhtar et al. [69], Silva et al. [70], Xin et al. [71], and Mingzheng et al. [72], each of which holds considerable acclaim and widespread application in the SA domain (see Table 2). In assessing the results of the model, standard performance metrics such as F-measure and G-mean are employed, recognized for their dependability in evaluating imbalanced data [73]. Notably, the proposed model surpassed all other models across all evaluative parameters, even outpacing the highest-performing model, Mingzheng et al. More specifically, the proposed approach cut the error rate by an impressive 32% and 25% in the F-measure and G-mean metrics respectively. Further scrutiny was applied by comparing the performance of the proposed model with a condensed version of the proposed approach, referred to as proposed without RL. This comparison uncovered that the fully developed model drastically curtailed the error rate by an estimated 51%. These compelling findings underscore the critical importance and effectiveness of the RL technique embedded in the proposed approach.

Fig. 5, which illustrates the receiver operating characteristic (ROC) curves for the methodologies listed in Table II, employs the area under the curve (AUC) as a crucial metric for assessing classifier efficiency. An AUC of 1 is indicative of impeccable differentiation capabilities, whereas a score of 0.5 equates to a performance level similar to random guessing. Impressively, the method we developed showcased exceptional proficiency, achieving an AUC of 0.61. This score not only reflects its ability to accurately distinguish between positive and negative outcomes but also reinforces the method's validity as an effective predictive instrument. Furthermore, the 'Proposed without RL' variant demonstrated substantial efficacy, achieving an AUC of 0.53. This performance confirms its capacity to accurately classify positive and negative cases. In contrast, the methodologies developed by Mingzheng et al. and Xin et al. recorded more modest AUC values of 0.50, demonstrating a performance level that does not quite reach the benchmark set by our proposed approach. This ROC analysis distinctly highlights the varying levels of effectiveness among the different methodologies examined.

The notable predictive strength of our proposed method, both independently and in conjunction with RL, is a testament to its robustness and reliability in predictive modeling. This success not only validates the approach but also suggests the possibility for future enhancements. The method's adaptability and high level of accuracy pave the way for its potential application in a broader range of predictive scenarios, offering promising prospects for advancements in the field of predictive analytics. This creates exciting opportunities for further research and development, potentially leading to even more refined and efficient predictive models in the future.

TABLE II. RESULTS OBTAINED USING THE PROPOSED MODEL AND OTHER STATE-OF-THE-ART MODELS FOR SA

Model	Accuracy	F-measure	G-means
Akhtar et al. [68]	0.695 ± 0.160	0.580 ± 0.041	0.695 ± 0.160
Akhtar et al. [69]	0.705 ± 0.015	0.580 ± 0.035	0.705 ± 0.015
Silva et al. [70]	0.825 ± 0.165	0.760 ± 0.263	0.825 ± 0.255
Xin et al. [71]	0.800 ± 0.015	0.700 ± 0.120	0.800 ± 0.000
Mingzheng et al. [72]	0.855 ± 0.105	0.820 ± 0.056	0.855 ± 0.269
Proposed without RL	0.860 ± 0.005	0.850 ± 0.012	0.860 ± 0.035
Proposed	0.891 ± 0.015	0.890 ± 0.003	0.898 ± 0.055

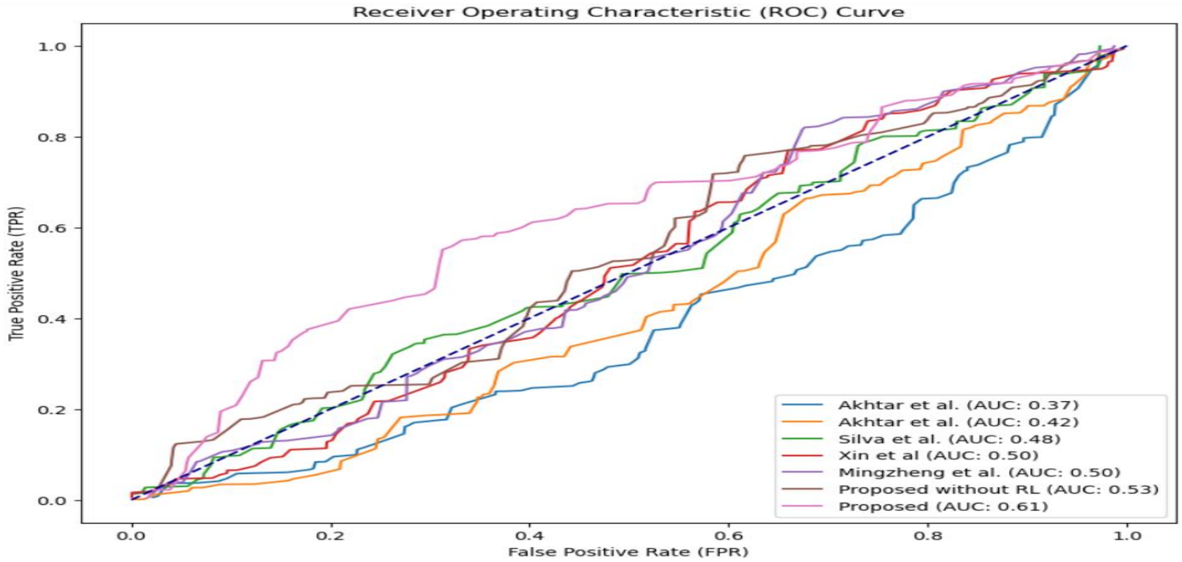


Fig. 5. ROC diagram for the proposed model and other methods. The Blue dashed line represents the ROC curve for a random guess.

2) *Impact of the reward function:* In SA, the allocation of rewards for both accurate and inaccurate classifications is distributed between majority and minority classes, marked as ±1 and ±λ, correspondingly. The determination of λ is closely associated with the ratio of majority to minority class instances. As this ratio increases, it is anticipated that the ideal λ value will demonstrate a corresponding decrease. To examine the effect of λ on model performance, the model underwent testing with a range of λ values, extending from 0 to 1 in increments of 0.1. During this process, the reward for the majority class remained constant. These experiments and their outcomes are clearly depicted in Fig. 6. When λ is set to 0, the significance afforded to the majority class is effectively negated, while a λ setting of 1 achieves an equilibrium in the treatment of both majority and minority classes. As illustrated in Fig. 6, the model's peak performance is observed at a λ value of 0.7, cutting across all evaluated metrics. This indicates that the most beneficial λ value does not lie at the extremes (0 or 1) but rather at a midpoint. It is crucial to recognize that while it's important to lessen the majority class's dominance by adjusting λ, selecting a value that is too low can adversely affect the model's overall effectiveness. The collected data underlines the significant impact of λ on the performance of the SA model. The ideal λ value is contingent upon the proportion of majority to minority samples, making its careful selection imperative for optimal results. This nuanced approach to λ selection is integral to balancing classification accuracy and ensuring the most advantageous outcomes in SA model applications.

3) *Impact of loss function:* In addressing the issue of skewed data distribution within SA, the utilization of various conventional methodologies, such as the adaptation of data augmentation techniques and loss functions, is essential. Of these methodologies, the selection of an apt loss function is

particularly significant as it can effectively accentuate the importance of the minority class. In this study, a rigorous evaluation of five distinct loss functions is conducted in this study, namely weighted cross-entropy (WCE) [74], balanced cross-entropy (BCE)] [75], dice loss (DL) [76], Tversky loss (TL) [77], and combo Loss (CL) [78]. These were assessed in relation to the proposed model. Notably, the BCE and WCE loss functions ensure unbiased consideration by assigning equivalent weightage to positive and negative samples. In this assortment of functions, the CL function has proven to be an efficient tool for applications wrestling with imbalanced data. This is achieved by its judicious weight distribution; wherein lesser weight is attributed to simpler examples and higher weight to complex ones. This strategic allocation emphasizes the challenging, often underrepresented cases without downplaying the simpler ones. As evidenced in Table III, the CL function, when compared to TL, generates a lower error rate, with the reduction ranging between 15% and 29% across accuracy and F-measure metrics. This remarkable reduction exhibits the CL function's competence in managing imbalanced data more effectively. However, notwithstanding these promising outcomes, it is vital to recognize that the performance of the CL function pales in comparison to RL, trailing behind by a significant 41% margin. This implies that while the CL function can be a feasible option for handling imbalanced data, more sophisticated techniques like RL may offer more precise results.

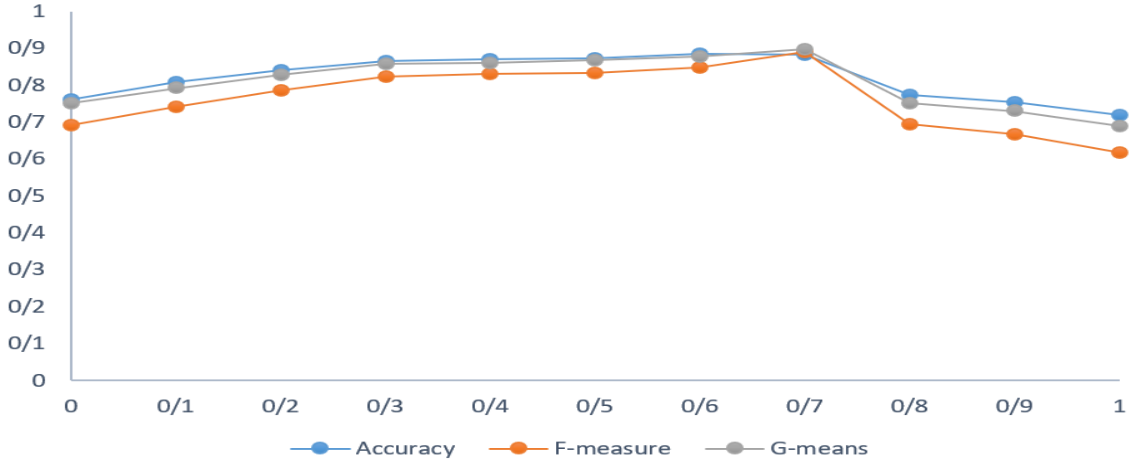


Fig. 6. The performance metrics of the suggested model were graphically represented in relation to the value of λ within the reward function.

TABLE III. RESULTS OF DIFFERENT LOSS FUNCTIONS FOR SA.

Loss function	Accuracy	F-measure	G-means
WCE	0.75±0.03	0.74±0.00	0.76±0.03
BCE	0.80±0.02	0.77±0.01	0.81±0.00
DL	0.81±0.03	0.80±0.01	0.82±0.00
TL	0.83±0.12	0.81±0.04	0.84±0.06
CL	0.86±0.00	0.84±0.04	0.86±0.15

4) Discussion: In the domain of big data and predictive analytics, this groundbreaking study introduces a pioneering approach that combines sentiment analysis from social media with historical stock prices to accurately forecast future stock prices. By harnessing the knowledge exchange platforms on the Internet, the authors offer a fresh perspective on stock market prediction, emphasizing the significant influence of public sentiment on financial market trends. The proposed research presents a novel methodology that analyzes social media by integrating public sentiment, opinions, news, and past stock prices to predict future stock prices. The methodology consists of two main stages. In the initial stage, a SA model is employed with three dilated convolution layers, enabling simultaneous feature extraction and classification. This approach effectively captures the underlying emotions expressed in user-generated content on social media, providing valuable insights into stock market trends. However, SA often faces the challenge of unbalanced classification, which occurs when one class of data is significantly more prevalent than the others. To overcome this challenge, an innovative RL strategy is introduced, treating the task as a sequential decision-making process. The agent receives rewards at each step for accurate classification, with smaller rewards assigned to the majority class compared to the minority class. This approach improves classification accuracy and enables the model to differentiate between different sentiments more effectively. In the second stage, the study incorporates an attention-based LSTM approach, which integrates historical stock prices and the sentiment analysis results obtained in the previous stage to predict future stock prices. The research validates the effectiveness of the proposed model through a series of ablation studies, confirming the positive contributions of the attention-based LSTM and RL components to the overall model performance. The innovative methodology represents a promising advancement in predictive analytics, particularly due to its unique integration of sentiment analysis and historical stock prices. Additionally, the RL strategy for addressing class imbalance adds an intriguing aspect to the field.

Fig. 7 provides a visual representation of the error trajectory observed in the proposed model. It demonstrates the model's progression over 150 training epochs, showcasing a consistent decrease in error. This signifies the continuous improvement of the model's ability to predict stock market trends. The gradual reduction in error rates indicates the convergence of the model towards an optimal solution. Notably, in the final stages of training, the error rates reach an exceptionally low level, such as 0.0000001. This significant decrease underscores the outstanding predictive accuracy achieved by LSTM.

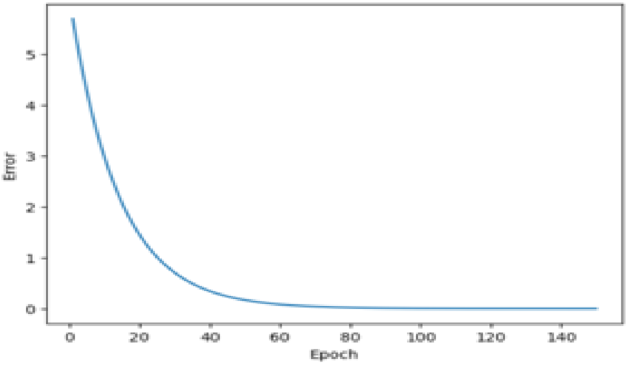


Fig. 7. Diagram of model error during 150 epochs.

Despite the encouraging findings obtained from the research, it is important to acknowledge and address certain limitations that emerged. These limitations not only provide opportunities for further investigation but also indicate areas where improvements can be made in future studies.

- The effectiveness of the proposed methodology relies on the quality and representativeness of the data utilized. However, social media posts and financial news, which serve as the primary data sources, can sometimes be unreliable or inaccurate, posing a risk to the analysis and prediction process [79]. To address this issue, future research can focus on refining data pre-processing techniques to filter out misleading or irrelevant posts, ensuring that sentiment analysis is based on accurate and reflective data [80]. Additionally, there is a need to develop more advanced NLP techniques that can better understand the context and sentiments expressed in social media posts and financial news. Expanding the scope of research, incorporating techniques like sentiment intensity analysis, emotion detection, irony and sarcasm detection, and stance detection can provide a more comprehensive understanding of public sentiment.
- Furthermore, exploring methods to assess the reliability and credibility of data sources, such as establishing a rating system for social media platforms or news outlets, can enhance the accuracy of sentiment analysis [81]. Looking ahead, the potential of machine learning and AI in this field is vast. Future work can explore the integration of other data forms, including audio and video content from different social media platforms, to gain more diverse insights into public sentiment and its impact on stock market trends [82].
- The current methodology employs a fixed reward ratio for the majority and minority classes, which has demonstrated effectiveness in the experiments. However, it may not be optimal for all datasets, particularly those with varying degrees of class imbalance. To address this, future research could explore the dynamic adjustment of the reward function based on the observed class distribution in the data.
- In addition, the proposed approach only considers immediate historical stock prices for prediction. To enhance the predictive capabilities of the model, future studies could investigate the inclusion of longer historical periods or incorporate other relevant factors such as industry trends, economic indicators, or global events. Furthermore, it is worth noting that the proposed method was tested on a limited number of stock markets. To ensure the generalizability of the model, it would be valuable to validate the approach using data from various global stock markets.
- A possible direction for future work could be to incorporate inductive learning elements, allowing the model to generalize from past trends and apply them to future predictions [83]. This extension would enable the model to capture and leverage the underlying patterns and dynamics of the stock market that span across different time periods. By integrating inductive learning, the model can learn from historical data and extract valuable insights that can be applied to forecast future stock prices with greater accuracy. This approach would involve incorporating techniques such as time-series analysis, trend identification, and pattern recognition to identify recurring patterns and trends in the data. By recognizing and understanding these patterns, the model can make more informed predictions about future market behavior [84].
- Additionally, incorporating inductive learning would help the model adapt and adjust its predictions as new data becomes available, ensuring its relevance and effectiveness in dynamic market environments. This approach holds great potential for improving the long-term forecasting capabilities of the model, enabling it to anticipate market trends and make proactive investment decisions. Overall, the incorporation of inductive learning elements represents an exciting avenue for future research, offering the opportunity to enhance the predictive power and robustness of the model in stock market analysis.

- The proposed model primarily focuses on analyzing sentiments from social media platforms. However, future work could look into incorporating various other types of online sources that influence the stock market. Such sources could include financial forums, professional analyst reports, business news websites, and investor behavior analytics from trading platforms [85]. These additional sources of data could provide a more comprehensive and diversified input into the model, leading to potentially higher predictive accuracy. Advanced web scraping techniques and API usage could be employed to harvest data from these varied sources, while machine learning algorithms can be applied to analyze and incorporate this data into the predictive model.

- The current SA model uses a single layer of sentiment classification, which categorizes the sentiment into positive, negative, or neutral. While this works effectively, it may overlook nuanced sentiment gradations that can significantly influence stock prices. Future research could introduce multi-dimensional sentiment analysis that captures not only the polarity of the sentiment but also the intensity and the emotional context. This can involve using techniques like aspect-based sentiment analysis (ABSA) [86], which extracts sentiments related to specific aspects of a topic. For example, a social media post might express positive sentiment about a company's management but a negative sentiment about its new product. Recognizing and capturing these nuanced sentiments can provide a more accurate sentiment representation and lead to improved stock price predictions.

- Considering the potential influence of influential individuals on the stock market, another future research direction could involve the analysis of sentiment expressed by key figures in the business world or financial analysts. Often, the views expressed by these individuals can significantly sway public sentiment and impact the stock market. Using entity recognition techniques, future models can differentiate comments made by these influential figures and assign higher weights to these in the sentiment analysis process.

- While the study provides an innovative approach to predicting stock prices, it is also important to study how the proposed methodology performs under various market conditions. Future work could include stress testing the model under different market scenarios such as bull markets, bear markets, and periods of high volatility. This would provide more insights into the model's robustness and reliability under different circumstances.

- Lastly, the use of deep learning methods in stock price prediction comes with the risk of overfitting, especially when the model becomes too complex [87]. Future research could look into the application of regularization techniques or ensemble methods to mitigate this risk. Additionally, more advanced model interpretability and explainability techniques can be explored to better understand the model's decision-making process and provide more insights to financial analysts and investors. This could further build confidence in the use of AI-based predictive models in financial markets.

V. Conclusion

This research introduces an innovative method for leveraging social media analysis to forecast future trends in stock prices. The proposed methodology encompasses two primary phases. During the initial phase, a SA model was constructed, employing three dilated convolutional layers. These layers were instrumental in simultaneously extracting feature vectors, which were subsequently amalgamated for the purpose of classification. Nonetheless, the SA model encountered a notable hurdle in the form of imbalanced classification. To address this challenge, a RL strategy was devised. This strategy framed the classification task as a series of sequential decision-making events, wherein the agent garnered rewards for each instance of precise classification. To effectively manage the issue of class disparity, the model was designed to allocate reduced rewards for classifications pertaining to the predominant class in contrast to those of the minority class. Additionally, an enhanced DE algorithm was applied for the initialization of the weights in the SA model. In the second phase of the methodology, an attention-based LSTM technique was employed. This phase intricately combined historical stock market data with insights derived from the SA model in the preceding stage. The integration of these two data sources was instrumental in generating more accurate predictions of future stock prices, showcasing the efficacy and potential of this dual-stage approach in the realm of financial forecasting.

The computational challenges inherent in large-scale models such as BERT are significant and widely recognized. While these models boast advanced capabilities, they are also associated with a considerable computational load, largely due to their architecture involving millions of parameters. To effectively address this issue, one promising strategy is the exploration and adoption of more streamlined and efficient variants of the BERT model. In this context, DistilBERT stands out as a leading example, epitomizing this innovative class of models. DistilBERT effectively preserves a majority of the functionality inherent in the original BERT model, while achieving a notable reduction in both the model's size and its computational requirements. This successful blend of high-level performance coupled with increased efficiency highlights the groundbreaking potential of such models. They offer a promising avenue for revolutionizing the computational framework within the realm of large-scale NLP, paving the way for more accessible and sustainable AI solutions in this domain.

References

[1] B. Sun and V. T. Ng, "Analyzing sentimental influence of posts on social networks," in Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2014: IEEE, pp. 546-551.

[2] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: a survey," International Journal, vol. 2, no. 6, pp. 282-292, 2012.

[3] F. Z. Xing, E. Cambria, and R. E. Welsch, "Intelligent asset allocation via market sentiment views," ieee ComputatioNal iNtelligeNCe magaziNe, vol. 13, no. 4, pp. 25-34, 2018.

[4] V. Barot, V. Kapadia, and S. Pandya, "QoS enabled IoT based low cost air quality monitoring system with power consumption optimization," Cybernetics and Information Technologies, vol. 20, no. 2, pp. 122-140, 2020.

[5] A. Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, and K. Kotecha, "Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease," PeerJ Computer Science, vol. 7, p. e369, 2021.

[6] M. Soleimani, Z. Forouzanfar, M. Soltani, and M. J. Harandi, "Imbalanced Multiclass Medical Data Classification based on Learning Automata and Neural Network," EAI Endorsed Transactions on AI and Robotics, vol. 2, 2023.

[7] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in International conference on intelligent computing, 2005: Springer, pp. 878-887.

[8] I. Mani and I. Zhang, "kNN approach to unbalanced data distributions: a case study involving information extraction," in Proceedings of workshop on learning from imbalanced datasets, 2003, vol. 126: ICML, pp. 1-7.

[9] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, Learning from imbalanced data sets. Springer, 2018.

[10] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, "Training deep neural networks on imbalanced data sets," in 2016 international joint conference on neural networks (IJCNN), 2016: IEEE, pp. 4368-4374.

[11] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 5375-5384.

[12] E. Lin, Q. Chen, and X. Qi, "Deep reinforcement learning for imbalanced classification," Applied Intelligence, vol. 50, pp. 2488-2502, 2020.

[13] S. V. Moravvej, M. Joodaki, M. J. M. Kahaki, and M. S. Sartakhti, "A method based on an attention mechanism to measure the similarity of two sentences," in 2021 7th International Conference on Web Research (ICWR), 2021: IEEE, pp. 238-242.

[14] S. V. Moravvej, A. Mirzaei, and M. Safayani, "Biomedical text summarization using conditional generative adversarial network (CGAN)," arXiv preprint arXiv:2110.11870, 2021.

[15] S. Moravvej, M. Maleki Kahaki, M. Salimi Sartakhti, and M. Joodaki, "Efficient GAN-based method for extractive summarization," Journal of Electrical and Computer Engineering Innovations (JECEI), vol. 10, no. 2, pp. 287-298, 2022.

[16] M. Marani, M. Soltani, M. Bahadori, M. Soleimani, and A. Moshayedi, "The Role of Biometric in Banking: A Review," EAI Endorsed Transactions on AI and Robotics, vol. 2, no. 1, 2023.

[17] S. V. Moravvej, S. J. Mousavirad, D. Oliva, and F. Mohammadi, "A Novel Plagiarism Detection Approach Combining BERT-based Word Embedding, Attention-based LSTMs and an Improved Differential Evolution Algorithm," arXiv preprint arXiv:2305.02374, 2023.

[18] H. Gharagozlou, J. Mohammadzadeh, A. Bastanfard, and S. S. Ghidary, "RLAS-BIABC: A reinforcement learning-based answer selection using the bert model boosted by an improved ABC algorithm," Computational Intelligence and Neuroscience, vol. 2022, 2022.

[19] T. Eltaeib and A. Mahmood, "Differential evolution: A survey and analysis," Applied Sciences, vol. 8, no. 10, p. 1945, 2018.

[20] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the artificial bee colony (ABC) algorithm in collaboration with the fog nodes in the Internet of Things three-layer architecture," in 2021 29th Iranian Conference on Electrical Engineering (ICEE), 2021: IEEE, pp. 509-513.

[21] N. Milosevic, "Equity forecast: Predicting long term stock price movement using machine learning," arXiv preprint arXiv:1603.00751, 2016.

[22] S. Pandya, A. Sur, and K. Kotecha, "Smart epidemic tunnel: IoT-based sensor-fusion assistive technology for COVID-19 disinfection," International Journal of Pervasive Computing and Communications, vol. 18, no. 4, pp. 376-387, 2022.

[23] A. Porshnev, I. Redkin, and A. Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis," in 2013 IEEE 13th International Conference on Data Mining Workshops, 2013: IEEE, pp. 440-444.

[24] M. Athale, S. Nakod, and A. Kumar, "Stock Analysis using Sentiment Analysis and Machine Learning," 2020.

[25] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies, practices and applications," International Journal of Scientific and Technology Research, vol. 9, no. 2, pp. 601-609, 2020.

[26] F. Z. Xing, E. Cambria, and R. E. Welsch, "Natural language based financial forecasting: a survey," Artificial Intelligence Review, vol. 50, no. 1, pp. 49-73, 2018.

[27] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Predictive sentiment analysis of tweets: A stock market application," in Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data: Third International Workshop, HCI-KDD 2013, Held at SouthCHI 2013, Maribor, Slovenia, July 1-3, 2013. Proceedings, 2013: Springer, pp. 77-88.

[28] D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, "Stock market predication using a linear regression," in 2017 international conference of electronics, communication and aerospace technology (ICECA), 2017, vol. 2: IEEE, pp. 510-513.

[29] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," Expert systems with applications, vol. 42, no. 1, pp. 259-268, 2015.

[30] C. I. Patel, D. Labana, S. Pandya, K. Modi, H. Ghayvat, and M. Awais, "Histogram of oriented gradient-based fusion of features for human action recognition in action video sequences," Sensors, vol. 20, no. 24, p. 7299, 2020.

[31] A. E. Khedr and N. Yaseen, "Predicting stock market behavior using data mining technique and news sentiment analysis," International Journal of Intelligent Systems and Applications, vol. 9, no. 7, p. 22, 2017.

[32] P. Li and A. Hawbani, "An efficient budget allocation algorithm for multi-channel advertising," in 2018 24th International Conference on Pattern Recognition (ICPR), 2018: IEEE, pp. 886-891.

[33] A. Carosia, G. P. Coelho, and A. Silva, "Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media," Applied Artificial Intelligence, vol. 34, no. 1, pp. 1-19, 2020.

[34] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers, 2014, pp. 69-78.

[35] A. Yoshihara, K. Fujikawa, K. Seki, and K. Uehara, "Predicting stock market trends by recurrent deep neural networks," in PRICAI 2014: Trends in Artificial Intelligence: 13th Pacific Rim International Conference on Artificial Intelligence, Gold Coast, QLD, Australia, December 1-5, 2014. Proceedings 13, 2014: Springer, pp. 759-769.

[36] W. Jiang, "Applications of deep learning in stock market prediction: recent progress," Expert Systems with Applications, vol. 184, p. 115537, 2021.

[37] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," The Journal of Supercomputing, vol. 76, pp. 2098-2118, 2020.

[38] A. Sun, M. Lachanski, and F. J. Fabozzi, "Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction," International Review of Financial Analysis, vol. 48, pp. 272-281, 2016.

[39] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," Journal of Ambient Intelligence and Humanized Computing, pp. 1-24, 2020.

[40] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," Neural Computing and Applications, vol. 32, pp. 1609-1628, 2020.

[41] M. Bahadori, M. Soltani, M. Soleimani, and M. Bahadori, "Statistical Modeling in Healthcare: Shaping the Future of Medical Research and Healthcare Delivery," in AI and IoT-Based Technologies for Precision Medicine: IGI Global, 2023, pp. 431-446.

[42] K. V. Price, "Differential evolution," Handbook of Optimization: From Classical to Modern Approach, pp. 187-214, 2013.

[43] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the cuckoo algorithm to optimizing the response time and energy consumption cost of fog nodes by considering collaboration in the fog layer," in 2021 5th International Conference on Internet of Things and Applications (IoT), 2021: IEEE, pp. 1-5.

[44] M. Arafa, E. A. Sallam, and M. Fahmy, "An enhanced differential evolution optimization algorithm," in 2014 fourth international conference on digital information and communication technology and its applications (DICTAP), 2014: IEEE, pp. 216-225.

[45] S. V. Moravvej, S. J. Mousavirad, D. Oliva, G. Schaefer, and Z. Sobhaninia, "An improved de algorithm to optimise the learning process of a bert-based plagiarism detection model," in 2022 IEEE Congress on Evolutionary Computation (CEC), 2022: IEEE, pp. 1-7.

[46] K. Deb, "A population-based algorithm-generator for real-parameter optimization," Soft Computing, vol. 9, pp. 236-253, 2005.

[47] S. V. Moravvej, M. J. M. Kahaki, M. S. Sartakhti, and A. Mirzaei, "A method based on attention mechanism using bidirectional long-short term memory (BLSTM) for question answering," in 2021 29th Iranian Conference on Electrical Engineering (ICEE), 2021: IEEE, pp. 460-464.

[48] M. S. Sartakhti, M. J. M. Kahaki, S. V. Moravvej, M. javadi Joortani, and A. Bagheri, "Persian language model based on BiLSTM model on COVID-19 corpus," in 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA), 2021: IEEE, pp. 1-5.

[49] S. V. Moravvej, S. J. Mousavirad, M. H. Moghadam, and M. Saadatmand, "An LSTM-based plagiarism detection via attention mechanism and a population-based approach for pre-training parameters with imbalanced classes," in Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28, 2021: Springer, pp. 690-701.

[50] L. Hong et al., "GAN ‐ LSTM ‐ 3D: An efficient method for lung tumour 3D reconstruction enhanced by attention ‐ based LSTM," CAAI Transactions on Intelligence Technology, 2023.

[51] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," in 2015 IEEE international conference on big data (big data), 2015: IEEE, pp. 2823-2824.

[52] H. Zareiamand, A. Darroudi, I. Mohammadi, S. V. Moravvej, S. Danaei, and R. Alizadehsani, "Cardiac Magnetic Resonance Imaging (CMRI) Applications in Patients with Chest Pain in the Emergency Department: A Narrative Review," Diagnostics, vol. 13, no. 16, p. 2667, 2023.

[53] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[54] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

[55] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," Journal of machine learning research, vol. 3, no. Aug, pp. 115-143, 2002.

[56] A. Graves, "Generating sequences with recurrent neural networks," arXiv preprint arXiv:1308.0850, 2013.

[57] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," arXiv preprint arXiv:1409.2329, 2014.

[58] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Computer Science, vol. 7, p. e623, 2021.

[59] S. V. Moravvej et al., "RLMD-PA: A reinforcement learning-based myocarditis diagnosis combined with a population-based algorithm for pretraining weights," Contrast Media & Molecular Imaging, vol. 2022, 2022.

[60] K. Zhang, G. Zhong, J. Dong, S. Wang, and Y. Wang, "Stock market prediction based on generative adversarial network," Procedia computer science, vol. 147, pp. 400-406, 2019.

[61] J. Rasheed, A. Jamil, A. A. Hameed, M. Ilyas, A. Özyavaş, and N. Ajlouni, "Improving stock prediction accuracy using cnn and lstm," in 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020: IEEE, pp. 1-5.

[62] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," Neural Computing and Applications, vol. 32, pp. 9713-9729, 2020.

[63] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock closing price prediction using machine learning techniques," Procedia computer science, vol. 167, pp. 599-606, 2020.

[64] P. Mehta, S. Pandya, and K. Kotecha, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," PeerJ Computer Science, vol. 7, p. e476, 2021.

[65] S. R. Riady, "Stock Price Prediction using Prophet Facebook Algorithm for BBCA and TLKM," International Journal of Advances in Data and Information Systems, vol. 4, no. 1, pp. 1-8, 2023.

[66] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, "Stock market index prediction using deep Transformer model," Expert Systems with Applications, vol. 208, p. 118128, 2022.

[67] S. BL and S. BR, "Combined deep learning classifiers for stock market prediction: integrating stock price and news sentiments," Kybernetes, vol. 52, no. 3, pp. 748-773, 2023.

[68] M. S. Akhtar, T. Garg, and A. Ekbal, "Multi-task learning for aspect term extraction and aspect sentiment classification," Neurocomputing, vol. 398, pp. 247-256, 2020.

[69] M. S. Akhtar, D. Gupta, A. Ekbal, and P. Bhattacharyya, "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis," Knowledge-Based Systems, vol. 125, pp. 116-135, 2017.

[70] J. Silva et al., "Algorithm for Detecting Opinion Polarity in Laptop and Restaurant Domains," Procedia Computer Science, vol. 170, pp. 977-982, 2020.

[71] X. Xin, A. Wumaier, Z. Kadeer, and J. He, "SSEMGAT: Syntactic and Semantic Enhanced Multi-Layer Graph Attention Network for Aspect-Level Sentiment Analysis," Applied Sciences, vol. 13, no. 8, p. 5085, 2023.

[72] L. Mingzheng, H. Zelin, L. Jiadong, and L. Wei, "Aspect-level sentiment analysis model fused with GPT and multi-layer attention," in Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022), 2023, vol. 12610: SPIE, pp. 279-284.

[73] S. Danaei et al., "Myocarditis Diagnosis: A Method using Mutual Learning-Based ABC and Reinforcement Learning," in 2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo), 2022: IEEE, pp. 000265-000270.

[74] Ö. Özdemir and E. B. Sönmez, "Weighted cross-entropy for unbalanced data with application on covid x-ray images," in 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020: IEEE, pp. 1-6.

[75] F. Huang, J. Li, and X. Zhu, "Balanced Symmetric Cross Entropy for Large Scale Imbalanced and Noisy Data," arXiv preprint arXiv:2007.01618, 2020.

[76] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced NLP tasks," arXiv preprint arXiv:1911.02855, 2019.

[77] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8, 2017: Springer, pp. 379-387.

[78] S. A. Taghanaki et al., "Combo loss: Handling input and output imbalance in multi-organ segmentation," Computerized Medical Imaging and Graphics, vol. 75, pp. 24-33, 2019.

[79] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD explorations newsletter, vol. 19, no. 1, pp. 22-36, 2017.

[80] B. Zhang, H. Yang, T. Zhou, M. Ali Babar, and X.-Y. Liu, "Enhancing financial sentiment analysis via retrieval augmented large language models," in Proceedings of the Fourth ACM International Conference on AI in Finance, 2023, pp. 349-356.

[81] X. Liu, C. Li, J. L. Nicolau, and M. Han, "The value of rating diversity within multidimensional rating system: Evidence from hotel booking platform," International Journal of Hospitality Management, vol. 110, p. 103434, 2023.

[82] K. Nyakurukwa and Y. Seetharam, "The evolution of studies on social media sentiment in the stock market: Insights from bibliometric analysis," Scientific African, p. e01596, 2023.

[83] A. Modirrousta-Galian, P. A. Higham, and T. Seabrooke, "Effects of inductive learning and gamification on news veracity discernment," Journal of Experimental Psychology: Applied, 2023.

[84] Z. Ding, J. Wu, Z. Li, Y. Ma, and V. Tresp, "Improving Few-Shot Inductive Learning on Temporal Knowledge Graphs using Confidence-Augmented Reinforcement Learning," arXiv preprint arXiv:2304.00613, 2023.

[85] H. Xia, S. Chen, J. Z. Zhang, and Y. Liu, "Superposition effect of online news on fintech platforms," International Journal of Emerging Markets, 2023.

[86] P. Mehra, "Unexpected surprise: Emotion analysis and aspect based sentiment analysis (ABSA) of user generated comments to study behavioral intentions of tourists," Tourism Management Perspectives, vol. 45, p. 101063, 2023.

[87] K. Olorunnimbe and H. Viktor, "Deep learning in the stock market—a systematic survey of practice, backtesting, and applications," Artificial Intelligence Review, vol. 56, no. 3, pp. 2057-2109, 2023.

Cloud Migration: Identifying the Sources of Potential Technical Challenges and Issues

Nevena Staevsky, Silvia Gaftandzhieva
University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

***Abstract*—Digital Transformation is emerging as a crucial factor for successful adaptation to the modern digital world for all possible economic and social entities. In recent years, cloud migration, cloud services and computing solutions adoption have been popular enablers for the Digital Transformation. During the Digital Transformation process, organizations and institutions face various technical challenges and implementation problems. This article explores the issues related to cloud migration and existing cloud service models. It investigates the advantages and disadvantages of the most popular cloud services offered by leading service providers, summarizes the main challenges in cloud migration processes, and how organizations can overcome them. Results help organizations understand the sources of potential technical challenges and implementation problems affecting cloud adoption and address these issues at an early stage of the initiative in order to reduce the threat of failure, avoid potential pitfalls and achieve desired cloud capabilities and business benefits.**

***Keywords*—Digital transformation; cloud; cloud migration; cloud models; PaaS; SaaS; IaaS; challenges**

I. Introduction

The topic of the research presented here is related to a field that until recently was considered evolving and nascent and currently occupies a central place in scientific analyses and practical developments. This is the area known as Digital Transformation, which is emerging as an important factor for successful adaptation to the modern digital world for all possible economic and social entities – countries, cities, industries, companies and people.

In fact, Digital Transformation began in the last century with the introduction of personal computers, the automation of industry, and the global expansion and establishment of the WWW as a means of communication and information exchange worldwide. As analogue systems were replaced by digital ones, the Internet became an integral part of personal and professional life.

The onset of this information age has given a strong impetus to worldwide globalisation, which is reaching hitherto unknown proportions.

The present study is part of a larger scientific research of the approaches to Digital Transformation of organizations and companies working in the field of the intangible sphere, such as, for example, service-oriented companies that provide online payment solutions, software or consulting services, educational and health organizations, as well as local or state government bodies, etc. The objective of the research is to identify typical implementation challenges and problems in the introduction of modern digital technologies. As a further matter, it aims to propose and experiment with solutions to overcome implementation problems faced by various organizations and institutions undergoing Digital Transformation.

According to Gartner [1], more than 85% of organizations will embrace a cloud-first principle by 2025. Since cloud migration and the adoption of cloud services and computing solutions are certainly enablers for Digital Transformation, the research should first address the challenges in this area and the approaches to optimize this multi-stage process.

This article presents the initial stages of this large-scale scientific research, which consists of exploring what cloud migration is and which cloud service models exist in cloud computing adoption. Furthermore, the paper investigates the advantages and disadvantages of the most popular cloud services offered by leading service providers and summarize the main challenges in cloud migration processes and how they can be optimized. The analysis performed allows us to understand the sources of potential technical challenges and implementation problems affecting cloud adoption, and to address these issues at an early stage of the initiative in order to reduce the threat of failure. The study serves to inform future efforts in implementing cloud innovation to avoid potential pitfalls and achieve desired cloud capabilities and business benefits.

The rest of the paper is organized as follows. Section II presents the benefits of cloud migration and discusses the advantages and disadvantages of the main types of cloud service models. Section III presents an overview of some of the most popular cloud computing platforms offered by the leading cloud service providers and discusses their advantages and disadvantages. Section IV summarises the migration challenges corresponding to different aspects and suggests potential approaches to how organizations and institutions can overcome or mitigate them. The Conclusion in Section V summarizes the contribution and limitations of this paper and plans for future work in the field is summarized in Section VI.

II. Cloud Migration. Cloud Computing Service Models

Cloud migration refers to the process of moving digital assets, including data, applications, workloads and IT processes, from on-premises or legacy infrastructure to a cloud-based environment or so-called cloud computing environment. This transition involves transferring computing resources, such as servers, storage, databases, networks and software applications, to a cloud service provider's infrastructure. The ultimate goal of cloud migration is to achieve greater scalability, flexibility, and cost-efficiency and enhanced performance while leveraging the benefits of cloud computing [2].

The significance of cloud migration in modern IT infrastructure stems from the transformative impact it can have on businesses, operations, and technological landscapes. The key reasons that drive industries to adopt cloud technologies are well known (see Fig. 1), but below the reasons why cloud migration is vital in today's IT environment are discussed in more detail.

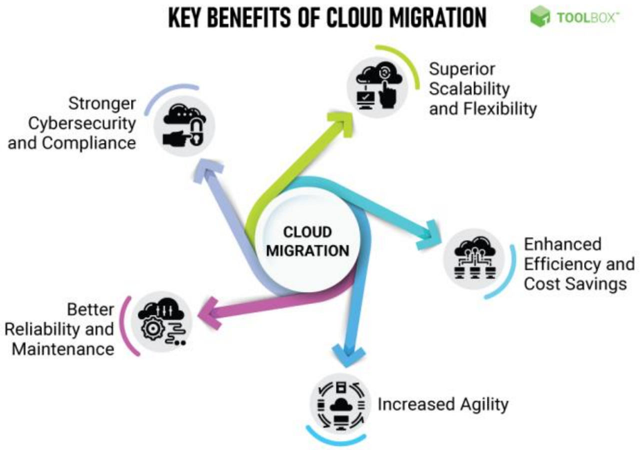

KEY BENEFITS OF CLOUD MIGRATION
TOOLBOX
CLOUD MIGRATION
Superior Scalability and Flexibility
Stronger Cybersecurity and Compliance
Enhanced Efficiency and Cost Savings
Better Reliability and Maintenance
Increased Agility

Fig. 1. Key benefits of cloud migration [3].

Cloud platforms offer elastic *scalability* [4], enabling businesses to easily scale resources up or down based on demand. This flexibility allows organizations to adapt to changes swiftly and efficiently without significant hardware investments or infrastructure reconfigurations.

Migration to the cloud often leads to cost savings [5] by moving from capital-intensive models (purchasing and maintaining physical hardware) to operational costs (pay-as-you-go or subscription-based models). This allows businesses to align costs with actual usage and optimize their IT spending.

Cloud migration fosters agility and innovation [6] by providing a platform for rapid development, testing, and deployment of new applications and features. Development teams can benefit from cloud-native services and tools, enabling faster time-to-market and a competitive edge.

Cloud infrastructure allows for worldwide accessibility to data and applications. Geographically distributed teams and users can access resources from any location, promoting collaboration and facilitating a globally connected workforce.

Cloud service providers invest heavily in security measures, often surpassing what many organizations can afford or manage independently. They offer advanced security features, compliance certifications, and regular updates to mitigate security risks and ensure compliance with industry-specific regulations [4].

Cloud providers offer robust disaster recovery and backup solutions, ensuring data integrity and availability in case of unforeseen events or system failures. Cloud-based disaster recovery strategies contribute to enhanced business continuity and reduced downtime.

Cloud platforms provide powerful analytics and machine learning capabilities, enabling organizations to derive actionable insights from their data through data analysis [6]. This data-driven approach can inform strategic decision-making and drive business growth.

Consolidating workloads in the cloud often leads to increased resource utilization and energy efficiency, contributing to a reduced overall carbon footprint compared to traditional on-premise data centers by environmentally sustainable solutions [7].

Migrating to the cloud allows companies to use centralized cloud services hosted by a vendor or cloud provider and delivered by the Internet. Cloud providers offer different cloud computing service models, each serving specific purposes and meeting different requirements. These service models are often called "as a service" or "aaS" models [8]. The three main types of service models are (see Fig. 2) software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS).

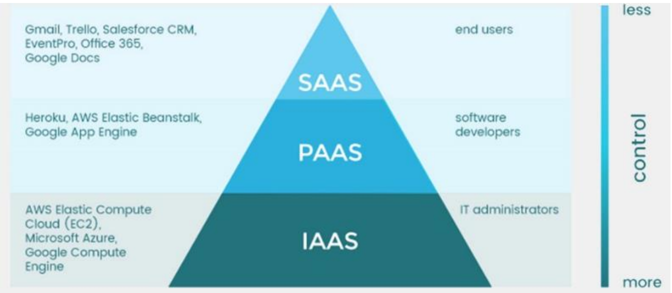

Gmail, Trello, Salesforce CRM, EventPro, Office 365, Google Docs
SAAS
end users
Heroku, AWS Elastic Beanstalk, Google App Engine
PAAS
software developers
AWS Elastic Compute Cloud (EC2), Microsoft Azure, Google Compute Engine
IAAS
IT administrators
less
control
more

Fig. 2. Main types of cloud computing service models [9].

Infrastructure as a Service (IaaS) offers virtualized computing resources over the Internet, including virtual machines, storage, networking, and other infrastructure components. Users can manage and control the entire infrastructure by installing and configuring operating systems, applications and databases. IaaS provides a highly flexible and customizable environment commonly used by IT administrators and developers to build test and manage their applications and services [10]. Examples of IaaS providers include Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), and IBM Cloud.

Platform as a Service (PaaS) provides a platform and environment for developers to create, deploy, and manage applications without worrying about the underlying infrastructure [11]. It offers development tools, frameworks, runtime environments, databases, and other services needed to develop applications [9]. Developers focus on coding and application logic, while the PaaS provider looks after the infrastructure management. Examples of PaaS are Microsoft Azure App Service, Google App Engine, and AWS Elastic Beanstalk.

Software as a Service (SaaS) delivers software applications over the Internet on a subscription basis. Users access these applications through web browsers without download or install anything locally [12]. The service provider manages all aspects of the application, including maintenance, security, updates and infrastructure. Examples of SaaS applications include email services (e.g., Gmail), customer relationship management (CRM) systems (e.g., Salesforce), and collaboration tools (e.g., Microsoft 365).

In practice, organisations often use a combination of these service models to meet their specific requirements, known as 'hybrid' or 'multi-cloud' approaches. This allows them to avoid disadvantages and leverage the strengths and advantages of each service model (see Table I) to create a customised solution that best suits their business needs.

TABLE I. CLOUD COMPUTING SERVICE MODEL (DIS)ADVANTAGES

Service Model	Advantages	Disadvantages
IaaS	Full control and customization of the infrastructure. Scalability and flexibility to meet changing needs. Pay-as-you-go pricing model that optimizes cost efficiency. Rapid provisioning and scaling of resources	Requires expertise in infrastructure management. Security, updates and maintenance are the responsibility of the user. Potential complexity in managing different infrastructure components
PaaS	Optimized application development and deployment. Automatic scaling depending on application demand. Built-in development tools and frameworks. Cost and time effective for development teams	Limited control over basic infrastructure. Dependence on the PaaS provider for updates and maintenance. Some constraints on the choice of development tools and frameworks
SaaS	No user installation or maintenance is required. Accessibility from any device with an internet connection. Regular automatic updates and patches. Scalability depending on user needs	Limited customization and control options compared to on-premises solutions. Dependence on the service provider for availability and security

In recent years, there has been a trend towards the formation of new "aaS" in the cloud, which are referred to under the general term Everything as a Service (XaaS). It is a broad term encompassing different service models where different types of resources and capabilities are provided over the Internet as a service. XaaS essentially extends the "as a service" concept beyond traditional cloud computing service models (SaaS, PaaS, IaaS) to include a wide range of offerings related to technology, software, infrastructure and even non-technology areas [13], as follows:

- Function as a Service (FaaS): FaaS is a serverless computing model that allows developers to execute code in response to events without managing servers. This is an example of XaaS in the context of application execution.

- Database as a Service (DBaaS): DBaaS provides database management and hosting in the cloud, allowing users to access and manage databases without the need to manage infrastructure.

- Security as a Service (SecaaS): This includes a set of security-related services, including threat detection, identity and access management, and security monitoring, provided in the cloud.

- Communication as a Service (CaaS): CaaS includes cloud-based communication services such as voice over IP (VoIP), video conferencing and messaging platforms.

- Monitoring as a Service (MaaS): MaaS provides application, infrastructure and network performance monitoring and monitoring services.

- Storage as a Service (StaaS): StaaS provides cloud-based storage solutions, including file storage, object storage, and backup services.

- IoT as a Service (IoTaaS): IoTaaS offers cloud-based services to manage and analyze data from Internet of Things (IoT) connected devices.

- Analytics as a Service (AaaS): AaaS provides cloud-based data analytics and processing capabilities, enabling organizations to make data-driven decisions to improve the efficiency of various aspects of the business, such as sales, demand forecasting, and sourcing.

- AI as a Service (AIaaS): AIaaS provides AI and machine learning services in the cloud, enabling organizations to leverage AI capabilities without the need for extensive AI expertise.

- Integration Platform as a Service (iPaaS): iPaaS is a cloud-based platform that facilitates integration between different applications, systems and data within an organization. iPaaS solutions enable seamless communication, data sharing, and workflow automation between disparate systems, both cloud and on-premises [14]. They play a critical role in simplifying the integration process, increasing business agility, and supporting digital transformation initiatives.

XaaS is a flexible and evolving concept, and new service models continue to emerge as technology evolves. It allows organizations to access and leverage a wide range of resources and capabilities without the need to make extensive investments in infrastructure or manage it, making it a key enabler of digital transformation and innovation across industries. According to research by Deloitte [15], companies adopting XaaS are increasingly aiming for agility and innovation rather than efficiency and cost reduction.

III. DISCUSSION: LEADING CLOUD SERVICE PLATFORMS - ADVANTAGES AND DISADVANTAGES OF POPULAR OPTIONS

Several leading cloud service providers have been dominating the cloud computing market over the past few years. Below is an overview of some of the most popular cloud computing platforms offered by the global players. Each of these options has its own set of features, addressing different needs and preferences. Still, they also share some common limitations that should be considered when deciding on the right cloud provider.

A. Amazon Web Services (AWS)

AWS offers a broad range of cloud services, including computing, storage, databases, machine learning, IoT, and more, providing flexibility for various use cases.

Not only does Amazon provide data centers in numerous regions worldwide, allowing for low-latency access to resources from different parts of the world [16], but also a wide range of security features and compliance certifications [17], making it suitable for industries with strict security requirements.

A variety of reliable storage solutions, including Amazon S3, Amazon EBS, Amazon Glacier, and others, offer scalable, durable, and highly affordable storage options for different use cases [17].

Furthermore, Amazon web services boast high performance and reliability [16], with strong achievement in terms of uptime and service availability, backed by Service Level Agreements (SLAs) to ensure reliability.

Due to the ability to scale resources up or down depending on demand, businesses can quickly adapt to changing requirements without upfront capital investment. Combined with that scalability and flexibility, the pay-as-you-go pricing model, which allows users to pay only for services they use, can help organizations manage and optimize costs effectively.

Also worth mentioning are the AWS's large and active user community, extensive documentation, and multiple third-party integrations in case customers need support and resources.

However, AWS has some weaknesses that should be mentioned. To start with, data transfer costs can add up and become a significant expense, especially for large volumes of data moved between regions or outside the AWS network. A further limitation of AWS is the lack of 24/7 support for all plans – only higher-level plans include around-the-clock access to customer support. Moreover, due to the huge customer base, custom support can be limited, and enterprises may experience delays in getting specialized help in critical situations. Finally, organizations that have invested in AWS may face challenges if they decide to migrate to another cloud provider due to potential vendor lock-in and the need for significant adjustments to existing systems and processes.

B. Microsoft Azure

Azure integrates seamlessly with Microsoft tools and products, making it an attractive choice for organizations already using Microsoft technologies such as Windows Server, Active Directory or SQL Server. Moreover, demonstrating a strong enterprise focus, the Microsoft cloud is also well suited for large enterprises, with a range of enterprise-class services, including Azure AD, Windows Virtual Desktop, and more.

Additionally, it provides solid support for hybrid cloud solutions [18], allowing organizations to easily connect on-premises data centers to the cloud.

Azure PaaS offerings could be considered fully comprehensive. It delivers a robust application development environment, including services such as Azure App Service and Azure Functions. Besides, it provides a rich set of developer tools and frameworks that support efficient application implementation and deployment.

Furthermore, Artificial intelligence (AI) and data analytics capabilities of Azure encompass advanced AI and machine learning (ML) services, such as Azure Machine Learning and Azure Cognitive Services, making it the preferred choice for AI-driven applications and data analytics.

Another significant service is the Azure Active Directory (Azure AD) with its robust identity and access management capabilities [18], enabling secure access to Azure services and other integrated applications.

Although the public cloud computing platform has many advantages, it also has a fair share of disadvantages, e.g. its complexity. The rich portfolio of services and features can be overwhelming for new users. Installing and initially configuring Azure services can get complex, especially when configuring networks, security, and permissions. Azure has a significant global presence. Even so, it has a slightly less global reach, compared to AWS in some parts of the world. Likewise, compared to AWS, Azure may have fewer third-party integrations and a more limited selection of third-party tools available in its ecosystem. Its cloud services possess limited platform portability because they are designed to run primarily within the Azure ecosystem, which can create challenges if an organization decides to move to another cloud service provider.

C. Google Cloud Platform (GCP)

GCP data analytics, ML and AI capabilities with services such as BigQuery, TensorFlow and Google AI Platform make it the preferred choice for organizations requiring advanced data processing and analytics.

A significant advantage of the platform is its powerful big data solutions such as BigQuery, Dataflow and Dataprep that enable organizations to efficiently process and analyze large data sets and gain valuable insights.

On the other hand, as a pioneer in containerization and Kubernetes, Google provides a robust Kubernetes Engine with its Cloud Platform. Accordingly, GCP [19] excellently suits companies focused on containerized applications and microservices.

As might be expected, GCP integrates seamlessly with Google Workspace (formerly G Suite), providing a complete environment for collaboration, productivity and cloud applications.

In addition, GCP has a massive global network infrastructure with strategically distributed data centers, which provides low-latency access to resources from different regions. Being designed for scalability and high performance, this infrastructure also allows businesses to scale resources quickly based on demand, ensuring optimal performance even as traffic spikes.

Due to its robust security features and regulatory compliance certifications, the platform ensures data privacy and meets various industry-specific regulatory requirements.

As for pricing, Google offers a pay-as-you-go pricing model that provides cost efficiencies and flexibility for businesses, especially those looking to manage and optimize costs.

Some disadvantages of GCP compared to AWS and Azure mentioned worth are the smaller market share and the limited global presence as it has fewer data center regions, which can impact latency and availability for some users. Despite being strong in certain areas, such as data analytics and machine learning, GCP may have a less diverse service portfolio than competitors such as AWS and Azure. Moreover, one may find the maturity of some services insufficient, especially of some relatively newer ones, in comparison with their AWS or Azure counterparts, which may result in occasional limitations or evolving features. A further weakness of Google cloud is its relatively limited focus on enterprises, although it is expanding its enterprise offerings. In the past GCP have been more focused on start-ups and developers, which may affect its attractiveness to larger enterprises. As for documentation and support [20], the platform's documentation is comprehensive, but support and specialized resources may not be as available as with larger cloud providers.

D. Alibaba Cloud

Alibaba Cloud has a significant presence in Asia and China, making it an excellent choice for businesses targeting this region.

This Cloud provides specialized solutions with a focus on e-commerce and retail enterprises, including artificial intelligence services for customer insights.

It is indisputable that Alibaba has diverse service offerings. The Cloud delivers a wide range of services, including computing, storage, database, big data, artificial intelligence, machine learning and IoT to meet a variety of business needs.

The company is also known for its cost-effectiveness and competitive pricing [21] compared to other major cloud providers, making it attractive to cost-conscious organizations.

Finally, Alibaba Cloud offers advanced distributed computing capabilities that enable high-performance computing, data processing and analytics at scale.

Alibaba Cloud was founded in 2009, only three years after AWS. Even so, it is a relatively new player in the cloud market and is characterised with some limitations. In the first place is its limited global reach. Although expanding, Alibaba Cloud has fewer data center regions outside of Asia [21], which impacts global availability. Secondly, English language support and documentation may not be as extensive as other providers. Furthermore, compared to other major cloud providers such as AWS, Azure and GCP, Alibaba Cloud's ecosystem may be less mature, with fewer third-party integrations and a smaller developer community. What can also affect its appeal to larger companies is its initially limited focus on enterprises, since Alibaba was more focused on start-ups and SMEs in the past. Another significant disadvantage concerns the compliance with international regulations and data protection and privacy that businesses outside China may have, especially given that Alibaba Cloud is a Chinese company. Finally, regarding the quality of customer support, some users report inconsistencies, having experienced problems with delays and getting timely help.

E. IBM Cloud

IBM Cloud has a strong focus on hybrid and multi-cloud services. It emphasizes hybrid and multi-cloud solutions [22] and is therefore an appropriate choice for organizations with complex infrastructure needs. Offering comprehensive support and consulting services, IBM helps organizations optimize their cloud infrastructure and ensure a successful transition to the cloud. As a matter of fact, IBM's Integration with Red Hat has strengthened the company's position in hybrid clouds with services such as IBM Cloud Paks and OpenShift.

Not only does the cloud offer a variety of enterprise-class services, including blockchain, artificial intelligence and IoT [22], but it simultaneously provides robust security features and regulatory compliance capabilities to ensure data security and privacy as well as compliance with numerous industry regulations and standards. The advanced AI and data analytics services [22], including IBM Watson, enable organizations to draw correct conclusions and make informed business decisions.

Furthermore, IBM Cloud has a global network of data centres, allowing companies to deploy resources worldwide and optimise performance by delivering resources closer to users.

However, IBM Cloud also possesses some shortcomings. It targets a niche market to meet more specific enterprise requirements and may not be as widely deployed as AWS, Azure or GCP. As a result, the cloud has a smaller market share compared to the major [23], which can lead to a smaller community, fewer third-party integrations and a narrower set of tools and services. In addition, some IBM services may be relatively newer or less mature than their AWS or Azure counterparts, which may result in evolving features or limitations. Finally, organizations that have invested in IBM technologies may face challenges if they decide to migrate to another cloud provider due to potential vendor lock-in and the need for significant adjustments to existing systems and processes.

F. Oracle Cloud

Oracle Cloud is known for its strong database solutions, making it an attractive choice for businesses relying heavily on Oracle databases.

It addresses the needs of enterprises with services such as Oracle Cloud Infrastructure, Oracle Autonomous Database and applications such as Oracle E-Business Suite. Oracle Cloud has an enterprise focus and provides a comprehensive suite of enterprise solutions, including databases, applications, computing, storage, analytics and more [24], suitable for large enterprises and businesses with complex needs.

Oracle Cloud integrates seamlessly with Oracle software and applications, providing a complete and optimized environment for organizations already using Oracle products.

Moreover, Oracle places a strong emphasis on security and regulatory compliance and offers a number of security features and certifications to ensure data protection and meet industry-specific regulatory requirements.

In addition, it has the capability to integrate with on-premises data centers, supporting a hybrid cloud model [25] for organizations that need a combination of cloud and on-premises solutions.

Still, some limitations should be pointed out regarding the cloud computing services offered by Oracle. Oracle Cloud's main strength lies in its niche market ─ database offerings. But it could be a weakness, as well, that may not allow it to meet all cloud computing needs. Oracle has a smaller number of data centres than the major cloud providers and, hence, a limited global reach. Besides, organizations that have invested heavily in Oracle technologies may face challenges if they decide to migrate to another cloud provider due to the potential vendor lock-in and the need for significant adjustments to existing systems and processes.

According to global data and business analytics platform Statista, the top three cloud providers by market distribution in 2023 are AWS, Azure and Google Cloud, followed at a significant distance by IBM and Oracle (see Fig. 3).

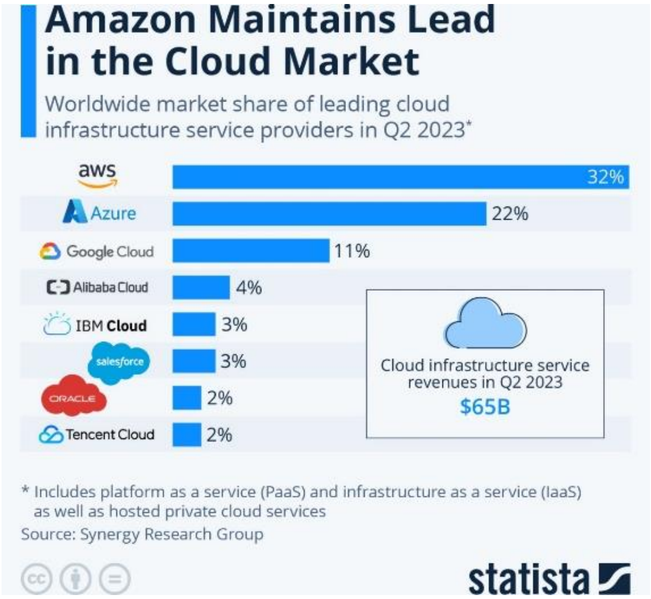


Fig. 3. Cloud market [26].

It is important to note that the evolution of cloud computing is extremely dynamic and these providers continue to improve and expand their services continuously. In addition, the specific needs and preferences of the individual organizations will determine which cloud provider is most suitable for them [26]. Therefore, a thorough evaluation and consideration of the advantages and disadvantages are critical when choosing a cloud service provider.

Beyond the provider-specific drawbacks mentioned above, there are some common ones that all cloud platforms share. When choosing to move to the cloud, organizations have to be aware of those limitations which could threaten their initiative, as for example – the steep learning curve. Because of the vast array of offered services and features, cloud platforms can be intimidating for users, especially for newbies or users unfamiliar with cloud computing concepts. Not only can this affect the ease of adoption and effective use of the platform, but it can also lead to extra training costs and unpredicted deployment delays.

A further consideration is the pricing complexity of the cloud services. Cloud providers offer different options for their customers in terms of pricing, but the most common models are reserved instances – purchased for a preliminary determined period, spot instances, where customers can bid for unused capacity, and the pay-as-you-go model, which charges customers based on the actual usage of resources. Even though the reserved instance option can lead to more significant cost savings, it has one major drawback – loss of flexibility [27]. On the other hand, the pay-as-you-go pricing model is flexible enough, but a more detailed examination of the granularity of the billing units proves drastic differences between the providers [27]. With the spot pricing, customers can get a significant discount, compared to the on-demand model, but there is always a risk of suddenly losing the purchased instances if the spot goes beyond your bid. Consequently, understanding and estimating costs can get very complex due to the multiple services delivered, the different pricing models and the various associated factors, such as compute resources, storage, data transfer, location, etc. [19].

Altogether, based on all above stated weaknesses of the different cloud service platforms offered, the following summarized technical problems and issues can be identified in the process of cloud migration and its further use:

- Resistance to deployment and employee leaving due to need for training to work with the cloud platform;
- Additional costs or respectfully incorrect forecasting of costs, caused by various factors - misunderstanding of pricing options; need for data transfer not planned in the initial budget; need for experienced experts for installation and configuration of cloud services; need for costs to implement missing third-party integrations; increased number of necessary services to satisfy business needs;
- Latency and delayed performance because of insufficient number of data center regions;
- Delay and temporary interruption of business activity, due to lack of expertise or timely customer support;
- Difficulties and need for additional costs for changes in existing systems and processes when migrating to another cloud service provider.
- Difficulties in performing daily operations caused by insufficient maturity of some services and their evolving.

IV. Results: Cloud Migration Challenges and Mitigation Approaches

Migrating existing on-premise applications to the cloud presents several key challenges that can be avoided through careful planning, appropriate strategies and effective implementation. Some of the key challenges, obstacles and difficulties that organizations face when transferring their infrastructure, applications, and data to cloud-based environments are related to a range of aspects concerning the migration process, such as data security and compliance, data transfer and latency, application compatibility, performance optimization, integration complexity, cost management, skill and knowledge gap, and downtime and business continuity.

Table II presents cloud migration challenges corresponding to the listed aspects and suggests potential approaches how to overcome or mitigate them.

TABLE II. Cloud Migration Challenges and Mitigation Opportunities

Integration aspect	Challenge	Mitigation
Data security and compliance:	Often on-premise applications handle sensitive data; hence, ensuring data security and regulatory compliance during migration is a serious concern.	Employ encryption, access controls, and compliance management tools to maintain data security. Conduct thorough compliance audits to ensure adherence to relevant regulatory requirements.
Data Transfer and Latency:	Transferring large volumes of data from on-premises systems to the cloud can be time-consuming and may cause latency issues.	Use data compression techniques, network optimization, and implement phased data transfer strategies to reduce the time and latency associated with data transfer [6].
Application compatibility:	Compatibility issues may arise when moving applications designed for on-premise environments to the cloud due to differences in infrastructure and configurations.	Conduct a comprehensive application assessment to identify compatibility issues. Modify or refactor the application as needed to ensure compatibility with the target cloud environment.
Performance optimization:	Ensuring optimal performance of applications in the cloud compared to on-premise can be challenging, affecting user experience.	Optimize applications for cloud infrastructure by utilizing auto-scaling, load balancing, and performance monitoring tools. Conduct performance testing to identify bottlenecks and optimize accordingly.
Integration complexity:	Integrating on-premise applications with cloud services and other applications can be complex, leading to data silos and inefficiencies.	Employ robust integration solutions and frameworks to facilitate seamless communication between on-premise and cloud-based components. Utilize APIs and middleware to bridge the integration gaps.
Cost management:	Cloud migration costs, including subscription fees and data transfer charges, can escalate and exceed the initial estimates.	Conduct a thorough cost analysis before migration, utilize cost management tools to monitor and control spending, and consider rightsizing resources based on usage patterns to optimize costs.
Skill and Knowledge Gap:	Migrating to the cloud often requires specialized skills and knowledge that may be lacking within the existing on-premises team.	Provide training and upskilling opportunities for the team or consider hiring cloud migration experts and consultants to bridge the skills gap and ensure a smooth migration process.
Downtime and Business Continuity:	Minimizing downtime during migration is critical to maintaining business continuity, which can be challenging during transition.	Implement a phased migration approach, perform thorough testing, and have a rollback plan in place to mitigate downtime and ensure uninterrupted business operations.

Addressing these challenges requires a holistic approach, involving thorough planning, stakeholder engagement, risk assessment, and leveraging appropriate technologies and methodologies to achieve a successful and smooth migration of on-premise applications to the cloud.

V. Conclusion

In summary, cloud migration is a strategic move that offers organizations the ability to align their IT infrastructure with business goals, enhance operational efficiency, optimize costs, and remain competitive in an ever-evolving technological landscape. By leveraging cloud capabilities, businesses can drive innovation, improve agility, and focus on their core competencies while harnessing the power of modern cloud computing.

The study presented in the paper is part of a wider scientific research with a very extensive goal to study and experiment with approaches, practices, problems and tools related to digital transformation initiatives in different spheres of the intangible domain.

This article draws on core challenges associated with cloud migration. By applying the proposed approaches for overcoming and mitigation, organizations and institutions can address these challenges at an early stage of the initiative and reduce the threat of failure. In this way, the study presented here can inform the future efforts of organizations and institutions in implementing cloud innovation to avoid potential pitfalls and achieve desired cloud capabilities and business benefits.

This study has some limitations as well. It is aimed at organizations and companies working in the intangible sphere, and all the challenges of cloud migration are brought out from this perspective. Therefore, additional research is needed to show whether all the proposed guidelines for overcoming or mitigating them can help avoid potential pitfalls and achieve desired cloud capabilities and business benefits in enterprises operating in the material sphere (e.g. production).

VI. Future Work

The results achieved here, related to the identification of the main sources of problems during the implementation of leading cloud service platforms, will allow throughout the subsequent stages of the research to systematize possible problems and implementation obstacles that may arise when moving data, applications, and other business elements in the process of cloud migration. As a result, best practices, recommendations, and solutions to overcome them will be proposed and experimented. Empirical evidence will be derived through experiments affecting individual stages at which business operations of specific business entities from the intangible sphere migrate to the cloud. Appropriate cloud services and a relevant cloud service provider will be selected for the investigation.

ACKNOWLEDGMENT

The paper is equally financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004-0001-C01 and the Scientific Research Fund at the University of Plovdiv "Paisii Hilendarski", project № MUPD23-FTF-019.

REFERENCES

[1] Gartner, “Gartner Says Cloud Will Be the Centerpiece of New Digital Experiences”, 2021, Accessed: Oct. 2023 at. https://www.gartner.com/en/newsroom/press-releases/2021-11-10-gartner-says-cloud-will-be-the-centerpiece-of-new-digital-experiences

[2] Z. Abbas, “Cloud Migration Strategies: Moving Applications and Workloads to the Cloud 2023”, Research Paper, 2023.

[3] H. Ashtari, “What Is Cloud Migration? Definition, Process, Benefits and Trends”, 2022, Accessed: Oct. 2023 at. https://www.spiceworks.com/tech/cloud/articles/what-is-cloud-migration/

[4] T. Guarda, F. Portela, M. F. Santos, “Advanced Research in Technologies, Information, Innovation and Sustainability: First International Conference”, ARTIIS 2021, La Libertad, Ecuador, November, Computer and Information Science Book 1485), Springer, ISBN: 978-3030902407, 2021.

[5] R. Amin, S. Vadlamudi, Md. M. Rahaman, “Opportunities and Challenges of Data Migration in Cloud”, Engineering International, 9(1), 2021, pp. 41-50, https://doi.org/10.18034/ei.v9i1.529

[6] J. A. Hernández, A. Hasayen, J, Aguado, Cloud Migration Handbook Vol. 1, Lulu Publishing Services, ISBN: 9781684709212, 2019, 278p.

[7] Accenture, “Cloud Migrations Can Reduce CO2 Emissions by Nearly 60 Million Tons a Year”, 2020, Accessed: Oct. 2023 at. https://newsroom.accenture.com/news/cloud-migrations-can-reduce-co2-emissions-by-nearly-60-million-tons-a-year-according-to-new-research-from-accenture.htm

[8] C. M. Mohammed, S. R. M. Zeebaree, S. R. M., “Sufficient Comparison Among CloudComputing Services: IaaS, PaaS, and SaaS: A Review”, International Journal of Science and Business, 5(2), 2021, pp. 17-30

[9] E. Plesky, “IaaS vs PaaS vs SaaS - cloud service models compared”, 2019, Accessed: Oct. 2023 at. https://www.plesk.com/blog/various/iaas-vs-paas-vs-saas-various-cloud-service-models-compared/

[10] A. Deldari, A. Salehan, A., “A survey on preemptible IaaS cloud instances: challenges, issues, opportunities, and advantages”. Iran J Comput Sci 4, 2021, pp. 1-24, https://doi.org/10.1007/s42044-020-00071-1

[11] M. Almubaddel, A. M. Elmogy, “Cloud computing antecedents, challenges, and directions”, Proceedings of the International Conference on Internet of Things and Cloud Computing, 2016, pp. 1-5

[12] S.K. Sowmya, “Layers of Cloud - IaaS, PaaS and SaaS: A Survey”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, pp. 4477-4480

[13] S. Bhattacharya, “XaaS: Everything-as-a-Service - The Lean and Agile Approach to Business Growth”, World Scientific Publishing Co Pte Ltd, 2021, https://doi.org/10.1142/11817

[14] Gartner, “Market Share Analysis: Integration Platform as a Service, Worldwide, 2022, Research”, 2023, Accessed: Oct. 2023 at. https://www.gartner.com/en/documents/4693099

[15] Deloitte, “Accelerating Agility with XaaS, A Report by the Center for Technology, Media and Telecommunications”, 2018

[16] J. Vliet, F. Paganelli, J. Geurtsen, “Resilience and Reliability on AWS: Engineering at Cloud Scale”, O’Reilly, ISBN 978-1-449-33919-7, 2018, 162 p.

[17] A. Bandaru, “Amazon Web Services”, 2020, Accessed: Oct. 2023 at. https://www.researchgate.net/publication/347442916_AMAZON_WEB_SERVICES

[18] C. Gill, S. Kuehn, “Configuring Windows Server Hybrid Advanced Services Exam Ref AZ-801: Configure advanced Windows Server services for on-premises, hybrid, and cloud environments”, Packt Publishing, ISBN: 978-1804615096, 2023, 602 p.

[19] J. J. Geewax, “Google Cloud Platform in Action”, Manning, ISBN: 9781638355908, 2018, 632 p.

[20] Collins Ayuya, “AWS vs Azure vs Google Cloud | Comparing Solutions 2023”, 2023, Accessed: Oct. 2023 at. https://www.channelinsider.com/cloud-computing/aws-vs-azure-vs-google-cloud/

[21] A. Dyachenko, “How does Alibaba Cloud compare to Microsoft Azure? 2021 Review”, 2021, Accessed: Oct. 2023 at. https://quickblox.com/blog/azure-vs-alibaba-cloud-platforms-comparison-2019/

[22] M. Law, “Top 10 biggest cloud providers in the world in 2023”, 2023, Accessed: Oct. 2023 at. https://technologymagazine.com/top10/top-10-biggest-cloud-providers-in-the-world-in-2023

[23] E. Jones, “Cloud Market Share: A Look at the Cloud Ecosystem”, 2023, Accessed: Oct. 2023 at. https://kinsta.com/blog/cloud-market-share/

[24] V. Arokia, “Oracle Cloud Services Vs. AWS: Which Is Better In 2023?”, Accessed: Oct. 2023 at. https://conneqtiongroup.com/blog/oracle-cloud-vs-aws-which-is-better

[25] F. Richer, “Amazon Maintains Lead in the Cloud Market”, 2023, Accessed: Oct. 2023 at. https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/

[26] M. Saraswat, R.C. Tripathi, “Cloud Computing: Comparison and Analysis of Cloud Service Providers-AWs, Microsoft and Google”, 9th International Conference System Modeling and Advancement in Research Trends (SMART), Electronic ISBN:978-1-7281-8908-6, 2020, pp. 281-285

[27] Sabrina Spilka, „Cloud Pricing Models - Shedding light upon pricing options”, 2021, Accessed: Oct. 2023 at https://www.exoscale.com/syslog/cloud-pricing-models/

Technology-Mediated Interventions for Autism Spectrum Disorder

Mihaela Chistol[1]*, Mirela Danubianu[2], Adina-Luminiţa Bărîlă[3]
Faculty of Electrical Engineering and Computer Science, Ştefan cel Mare University, Suceava, Romania[1, 2, 3]
ASSIST Software SRL, Suceava, Romania[1]

***Abstract*—According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), Autism Spectrum Disorder (ASD) is a complex neurological and developmental condition characterized by impairments in social interaction and communication. Despite significant advancements in the research field, no pharmaceutical medication has been designed for ASD treatment. Therefore, ASD treatment relies mainly on therapeutic intervention. Interactive technologies have emerged as valuable therapy augmentation tools. This research focuses on interactive technologies developed for ASD therapeutic intervention. The study introduces a conceptual framework for understanding the full spectrum of technologies involved in the ASD context. The employed methodology encompasses expert opinions and entails a cross-sectional study that included 59 participants with significant experience in interacting with individuals diagnosed with ASD in various real-life settings, including therapists, teachers, and parents of children with ASD. The research findings revealed a broad spectrum of technologies involved in ASD interventions, including applications, devices, and robots. The results bring a new perspective on the interactive technologies used in the therapy and diagnosis of ASD and highlight their important characteristics that can serve as a standard in the development of future technological solutions.**

***Keywords*—Autism spectrum disorder; technology-mediated interventions; assistive technologies; therapy; cross-sectional study**

I. Introduction

The technological prowess of human beings has been a defining characteristic since the dawn of our existence. Even in ancient times, our ancestors created innovative tools from stone to help them survive and work more efficiently. As time has passed, our desire to augment our lives through technology has only increased, and it is now a ubiquitous presence in all aspects of modern society. We rely on technology to communicate, enhance our physical capabilities, perform complex computations, and even cure medical conditions. From the earliest inventions to the latest cutting-edge breakthroughs, technology has been the driving force behind human progress and the key to unlocking new levels of knowledge and achievement. Technology is playing a critical role in shaping our world and advancing civilizations to new heights. But sometimes, some peaks are difficult to conquer. Similar to climbing Mount Everest, the development of medical technologies is a challenge because it is difficult to create applications and devices that efficiently diagnose and treat a disease. In the last century, Leo Kanner published a groundbreaking paper about early infantile autism [1] initiating a quest among therapists, doctors, and researchers to devise effective practices and interventions for treating autism spectrum disorder (ASD). However, it is a difficult task as this human condition ASD is a neurodevelopmental disorder characterized by deficits in social communication and the presence of restricted interests and repetitive behaviors [2]. Autism is frequently accompanied by other conditions such as epilepsy, depression, anxiety, attention deficit and hyperactivity disorder, as well as difficult behaviors like self-injury and sleep disturbances [3]. As a result, some individuals with autism can function independently, while others require lifelong assistance due to the severity of their disabilities.

Despite the fact that contemporary research shows people with autism are highly interested in using technology, they often require support from an intermediary person such as a caregiver, to use it effectively. This research enlisted the help of parents, teachers, and therapists who had children with ASD under their care. Their insights into the technologies used, goals pursued, desirable features, and drawbacks, as well as the digital solutions' adaptability to the social-cultural context, were instrumental in identifying the interactive technology aimed at ASD therapy.

According to the European Parliament's Research Service technologies for ASDs are less mature [4] and effects of technology-mediated interventions (TMIs) in ASD should be better understood. Thus, we adopted a new methodological strategy to understand the usability of technology-mediated therapeutic interventions in ASD. This approach involved the collection of expert opinions, real-world experiences of caregivers, and academic research. This innovative framework added depth and clarity to the results obtained. The structure of this paper is designed to provide a comprehensive understanding of the benefits and limitations of using interactive technologies in autism therapy. Section II of this paper presents the methodology. Section III offers a detailed analysis of the results and findings. In Section IV, we delve into a discussion about limitations of this study. Finally, we conclude the paper in Section V by summarizing the key takeaways and implications of our findings.

II. Materials and Methods

The main objective of this research is to explore the most efficient ASD therapies and cutting-edge technologies that can be integrated into therapy to facilitate the treatment of children diagnosed with autism. To address each topic of interest, we formulated three research questions presented in Table I.

*Corresponding Author.

TABLE I. RESEARCH QUESTIONS

ID	Research Question (RQ)
RQ_1	What therapies are used in the ASD treatment?
RQ_2	What technologies are employed in ASD therapeutic intervention?
RQ_3	What characteristics should technologies employed in ASD therapeutic interventions possess in order to effectively meet the needs of individuals with autism and promote positive outcomes?

In this study, we employed two research methods: expert opinion and cross-sectional study. Expert opinion was used to collect and analyze the opinions of professionals who have extensive experience working with individuals with autism, including medical practitioners and therapists. Through in-depth interviews, we were able to gain a deep understanding of the most effective therapeutic strategies. Cross-sectional study was conducted to explore and collect data directly from the end-users of the ASD technologies, including parents, therapists and teachers. This allowed us to obtain feedback on the usability and effectiveness of the interactive technologies in real-world settings. The research methodology was approved by the Research Ethics Board of Ștefan cel Mare University of Suceava, Romania, approval number 128.

A. Participants

1) Expert opinion participants: Expert opinion participants were selected based on rigorous criteria to ensure their qualifications. These criteria included: having accredited studies, approved by the College of Psychologists in Romania, possess extensive professional experience, showcase the ability to develop effective therapeutic intervention plans, and actively engage in academic activities such as participation in conferences, workshops, and research studies. As a result of selection process, two experts were invited to collaborate. The first expert brings more than a decade of experience in the field of psychology. He holds a certificate of practice in clinical psychology and specializes in Ericksonian Psychotherapy and Hypnosis. Additionally, he is internationally accredited as a Board Certified Behavior Analyst (BCBA). The second expert has over nine years of experience in Applied Behavior Analysis (ABA) therapy and is internationally accredited as a BCBA in behavioral analysis.

2) Cross-sectional study participants: Cross-sectional study involved individuals who have regular interactions with children diagnosed with ASD including: therapists, teachers, and parents. Over 100 individuals affiliated with the Association for Autism Intervention Suceava (AIAS), located in Romania, were invited to participate in the study and share their experiences by completing the survey using the web-based version of Google Forms. A total of 60 participants responded to the survey.

Fig. 1 illustrates the distribution of participants based on their roles, indicating that the majority of the participants 74.6% were parents of children with autism, while therapists accounted 23.7%. Unfortunately, only one teacher expressed interest in participating in the study, highlighting the educational system's reluctance towards individuals diagnosed with autism in Romania.

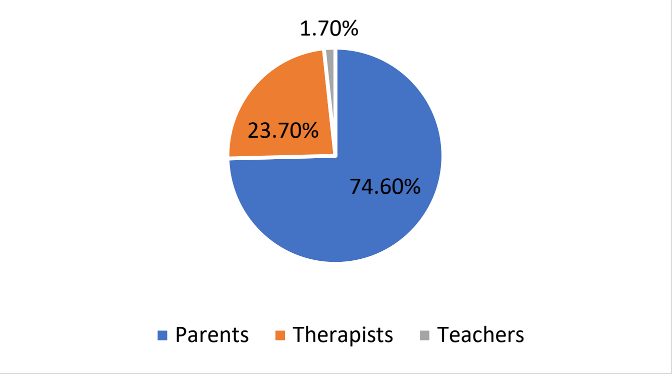


Fig. 1. Distribution of participants in the cross-sectional study based on their role in the lives of patients with autism.

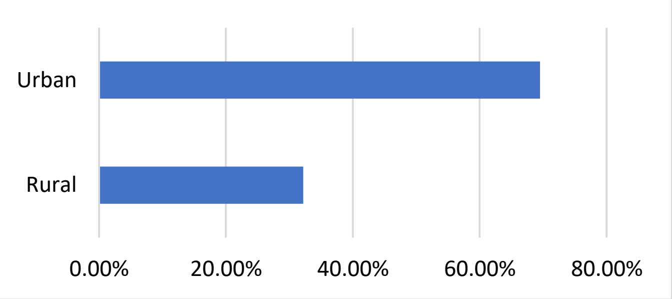


Fig. 2. Administrative organization in which the study participants reside.

In order to capture the diversity of personal experiences and socioeconomic backgrounds, the study included individuals from urban and rural administrative organizations. Fig. 2 presents the distribution of study participants based on their residence. The data reveals that the proportion of participants living in rural areas was 32.2%, which is lower than the proportion from urban areas 67.8%.

The importance of the participants' experience in interacting with people affected by ASD was the central aspect. Thus, we included both individuals with extensive experience of over 18 years, as well as beginners with only 1 year of experience in order to ensure diversity in the study. This method allowed to highlight both traditional and innovative approaches. Fig. 3 illustrates the distribution of participants' experience in years.

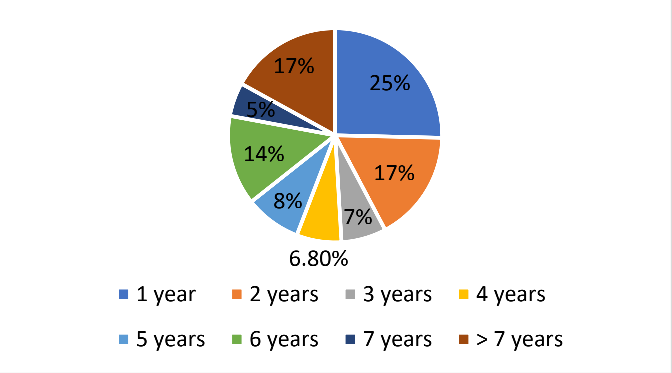


Fig. 3. Participants' experience expressed in years of interaction with ASD patients.

More than 22.1% of the participants are highly experienced individuals who possess a deep understanding of the behavior of patients with ASD. Meanwhile, 25.4% of the participants are novices with just 1 year of experience, which makes them well-suited for analyzing diagnostic methods, as their experiences are fresh. Additionally, 52.5% of participants possess notable experience ranging from two to six years.

B. Data Collection

1) Expert opinion data collection: Collaboration with therapists began in November 2022 and lasted six months. This partnership involved a mix of physical and online meetings, using the Microsoft Teams platform, version 1.0. The team interacted with therapists to discuss diverse aspects of therapeutic methods, the use of technology, and the specific abilities and disabilities of children with ASD.

The collaboration with therapists went beyond interviews, as we had the opportunity to observe therapy sessions in a real environment at the AIAS therapy center (see Fig. 4). During our observation, we paid attention to the educational materials used in therapy, including their content and visual representation. Furthermore, we became aware of the importance of the input type available in technologies for ASD, as we noticed that many children had challenges such as fine motor disabilities or speech impairments.

2) Cross-sectional study data collection: Conducting a cross-sectional study is a common method for collecting data from a specific population at a particular time. Our study focused on children diagnosed with ASD and their caregivers, with Romania serving as our research location. To reach our target population, 100 individuals, affiliated with AIAS therapy center, were invited via emails and social media to participate in the research, of which 59 expressed their willingness to participate, indicating a response rate of 59%. Ethical standards where followed during the study, the researchers informed participants about the study's scope and purpose, and participants were required to sign an agreement before their involvement.

To collect data, we used a survey questionnaire designed to explore the utilization of technologies in autism context. The questionnaire was created using Google Forms. Google Forms is a web-based app developed by Google which is used to create forms for data collection purposes [5]. The questionnaire was made available on February 1, 2023, and the participants' responses were collected over a period of three weeks.

The survey contained open-ended questions, multiple-choice questions, and Likert scales to evaluate the importance of certain factors related to technology use for ASD diagnosis, therapy, and entertainment. The survey consisted of four distinct categories. The first category focused on the experience of parents and therapists, posing questions about their background and experience in caring for children with autism. The second category inquired about the child's diagnosis and abilities, asking about the age of diagnosis and the child's functional level. The third category was centered around technologies and personal experiences in using them, with participants answering questions about their use of different technologies in the autism context.

Fig. 4. The collaboration with the ASD experts from the AIAS therapy center located in Suceava, Romania.

The fourth category was directed at identifying the best characteristics and features of technologies suitable for children with autism.

After concluding the data collection phase, we proceeded to perform statistical analysis on the gathered responses in order to detect patterns and trends within the data. The survey questionnaire results offered valuable insights into the use of technologies and therapeutic interventions for children with autism in Romania. These findings provided an overview of the real needs of parents, caregivers and children with ASD, highlighting the challenges they face, especially when access to formal therapy services is limited due to financial constraints.

III. Results

The results of this study are structured into three subsections, each addressing a specific research question:

Subsection A "Therapies for ASD Treatment" focuses on research question RQ_1 and reviews therapies that have demonstrated efficacy in ameliorating ASD symptoms.

Subsection B "Technologies in ASD" addresses research question RQ_2 and presents the technologies used to improve outcomes and provide support for people with ASD.

Subsection C "Characteristics of Technologies in ASD Therapeutic Interventions" answers RQ_3 and presents the key features of therapeutic technologies.

The results from the expert opinion and cross-sectional study are presented in a cohesive manner. By coupling theoretical frameworks and empirical evidence we aim to provide new insights and implications in the field of technology-mediated ASD interventions.

A. Therapies for ASD Treatment

The World Health Organization (WHO) estimates that worldwide about 1 in 100 children has autism [3]. The statistics surrounding this condition have raised concerns and encouraged researchers to investigate its etymology and potential treatments. Advances in autism research have gone hand in hand with significant progress in international policy. In May 2014, the Sixty-seventh World Health Assembly adopted a resolution entitled "Comprehensive and coordinated efforts for the management of autism spectrum disorders", which was supported by more than 60 countries. The resolution urges WHO to collaborate with Member States and partner agencies to strengthen national capacities to address ASD and other developmental disabilities ("Autism," n.d.). These legislative, social, medical and research efforts have led to the development of intervention and support strategies for people with ASD. In many countries, children can benefit from therapy and special education starting with kindergarten. The education learning plan is developed by a team of professionals and the child's parent. It is very important that parents are involved in the decisions that affect the education of their child. During the learning process, many types of therapies are applied. The most effective therapies according to the experts opinion are presented below.

1) Behavioral therapy: ABA is the most well supported intervention for ASD. ABA is the use of scientifically based behavioral principles in everyday situations. ABA Therapy works toward goals that help to increase or decrease different behaviors. All ABA programs share similar components, including specialized instructional strategies and parental involvement. ABA helps teaching skills that can be used at home, school, and in other settings [6]. Most commonly used ABA programs include: early intensive behavioral intervention (EIBI), positive behavioral and support (PBS), pivotal response training (PRT), discrete trial teaching (DTT) and relationship development interventions (RDI) [7].

2) Communication, speech and language therapy: Individuals diagnosed with ASD can experience communication difficulties regardless of their condition. Some individuals may not have acquired verbal communication skills at all, while others may have strong verbal abilities but struggle with social communication and interaction.

Speech and language therapy can help people with autism to improve their abilities to communicate and interact with others [8]. The therapy is designed to improve communication skills for both verbal and non-verbal individuals. It focuses on developing verbal communication skills, including pronunciation, syntax, and grammar, to help the person make themselves understood more easily and on developing non-verbal communication skills such as eye contact, gestures and body language, as well as on learning to interpret and respond to social cues.

3) Occupational therapy: Occupational therapy (OT) helps people ASD do everyday tasks by finding ways to work within and make the most of their needs, abilities, and interests [9]. Occupational therapists contribute to the care of children and adults with intellectual and developmental disabilities by focusing on activities and goals that are meaningful to the individuals and their families. Relevant performance areas include cognitive, sensory, perceptual, motor and psychosocial [10]. Occupational therapy methods include play-based activities as a way to help patients develop social and communication skills, and learning through movement activities to help patients improve motor and coordination aptitudes. Assistive technologies (ATs) are another important tool used in occupational therapy. Communication through tablets or other devices can help individuals with ASD to develop their communication skills and express their thoughts and emotions. .

4) Physical therapy: The way a person can use their motor skills impacts the ability to perform tasks. Sometimes individuals with autism have less developed motor skills and completing simple activities such as walking, climbing, or dressing is a challenge for them. Physical therapy can help in treating these disabilities. Physical therapy methods for individuals with autism may include adapted physical training activities and games that involve movement. In physical therapy the therapist works on physical limitations to help a person to develop muscles, balance and coordination needed for day-to-day activities [6]. The effectiveness of therapy in treating individuals with autism can vary depending on the severity of the condition.

5) Cognitive behavior therapy: Cognitive Behavioral Therapy (CBT) refers to a group of well-researched techniques that are effective in treating difficulties experienced by children and adults. CBT works well for treating anxiety and mood disorders, teaching stress and anger management, and improving interpersonal skills [6]. CBT generally consists of 12 to 18 1-h sessions and focuses on identifying and changing problematic thinking and behavioral patterns that maintain the youth's presenting symptomatology [11]. CBT combines two different approaches: cognitive therapy and behavioral therapy, to help change maladaptive thoughts and behaviors. Cognitive therapy focuses on changing the negative or distorted thoughts and perceptions of people with autism by learning techniques for self-observation and reflection on thoughts and feelings, identifying and replacing negative thoughts with more positive and realistic thoughts. Behavioral therapy focuses on changing maladaptive behaviors by learning new and healthier behaviors, such as developing social skills or developing a stress management program. CBT techniques have proven their effectiveness in studies conducted on adults and adolescents, clinical experience shows that individual therapy with children of preschool/school age must involve a certain level of play therapy.

Svetha Venkatesh from Curtin University, Australia et al. presented the “Playpad” multimedia-based system for delivering early intervention therapy for autism [12]. The system is based on cognitive therapy and allows individuals with autism to learn by performing natural interactions such as pointing and touching. Parents or non-experts can use the system to deliver home-based therapy, while therapists can construct lessons by specifying stimulus concepts and variations. Trials of the system have shown its effectiveness in training both adults and autistic children.

6) Art therapy: ASD is a condition that can affect sensorial functions, communication, and interpersonal relationships, leading to challenges in areas such as emotional, social, and behavioral interactions. Art therapy has the potential to address these complex issues due to its multisensory nature and relational approach [13]. Art therapy allows people with ASD to use their already visually-minded brains to communicate through artistic media. They can record images and visual data, express ideas and process memories that they are unable to do verbally [14]. Engaging in art therapy on a regular basis can benefit ASD individuals by promoting better family interactions, enhancing self-esteem, and improving emotional regulation both at school and home. Exposure to a variety of art materials can also contribute to the development of fine and gross motor dexterities, while providing children with the ability to adapt easily to new or unfamiliar situations. Most art therapy sessions are one-on-one, they can occur in a group setting, too. Working collaboratively on a single piece of art also fosters peer relationships. In this type of setting, one child draws something and the picture is passed to the next person, who adds to the work until everyone has contributed their part. This activity allows the children to acknowledge those around them and be more aware of others and their involvement in the project.

7) Autism therapies in practice: There is no prescription medication designed to treat ASD [6]. For this reason, therapies become extremely important in the treatment process of this condition. Choosing the appropriate therapy to treat the individual with ASD is a complex process that involves a multitude of factors, including patient symptoms, comorbidities, family, and physician expertise. In addition to these obvious factors, the choice of therapeutic method is also influenced by social norms, cultural values, and legal regulations. In countries that have progressive health insurance legislation, there is a significant evolution in the diversity of medical interventions for ASD treatment. However, in countries with a less developed medical system such as Romania, the diversity of therapies can be limited. In Romania the education for children with special needs began with segregation during the communist period [15]. In this context, our study aimed to identify the therapies practiced in real-life settings for the treatment of ASD symptoms in Romania. To achieve this, we conducted a cross-sectional study that involved 59 participants including parents and therapists who have experience in working with children with autism. Given the small data size we performed the statistical analysis of the answers received and identified that 84.7% of the study participants currently use therapeutic methods to improve children’s abilities. A worrying fact is that more than 6% of the participants did not use any therapy at all, despite the ASD diagnosis (see Fig. 5). Experts in the field suggest that the abandonment of therapy practice can be attributed to insufficient financial resources.

Families often depend solely on the income from the patient’s personal caregiver role. Furthermore, the limited availability of specialized facilities and services in Romania results in difficulties in accessing therapeutic support. The geographical distance between family residences and therapeutic centers presents another challenge, making it harder for individuals with ASD to access treatments.

Fig. 6 presents the therapies practiced in real life in Romania. We identified that the most commonly used therapies in the treatment of ASD patients are classic therapies, such as ABA Therapy, communication, speech and language therapy, sensory integration, and physical therapy. Many of these therapies are performed in an integrative way, combined with other complementary therapies, such as art therapy and 3C therapy.

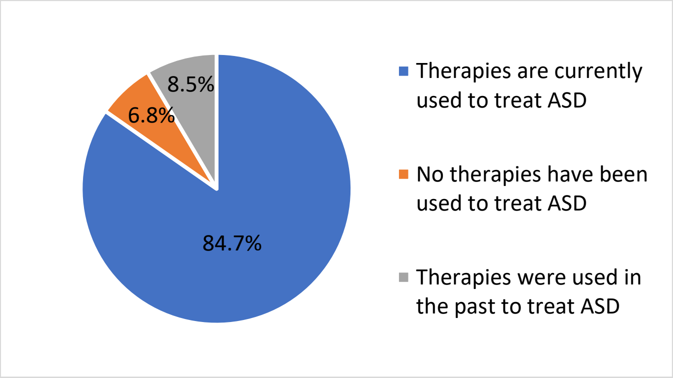


Fig. 5. Statistics on the use of ASD therapies.

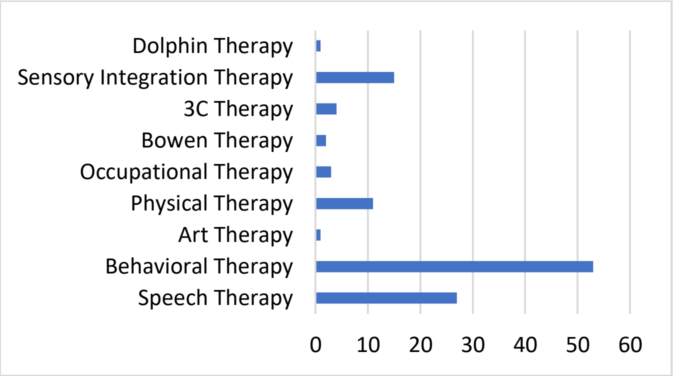


Fig. 6. ASD therapies practiced in real-life by the study participants.

B. *Technologies in ASD*

The medical field has evolved significantly in the 21st century due to the strong involvement of new technologies in diagnosis and treatment methods. In contemporary society, concepts such as telehealth and telemedicine have become common. Technologies are used for providing different medical services such as medical consultations through video calls, real-time diagnosis, or remote monitoring despite the fact that patients and medical professionals are in different locations. ASD therapy has also been impacted by technological evolution. Therapeutic strategies have integrated various types of interactive technologies to improve patients' abilities.

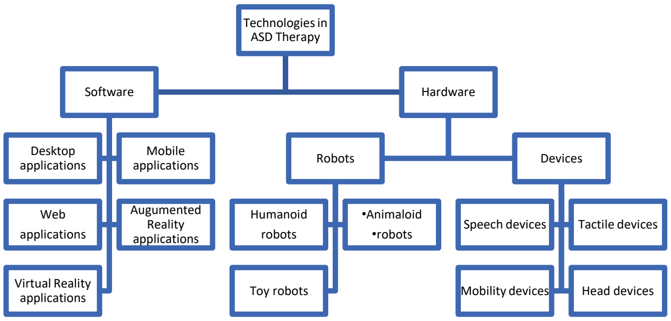


Fig. 7. Classification of technologies used in ASD therapy.

Studies [16], [17], [18] highlighted the idea that technologies used in teaching and therapy are well accepted by individuals with autism.

Technology can be defined as any electronic item, equipment, application, or virtual network that is used intentionally to increase, maintain, or improve daily living, work, productivity, recreation, and leisure capabilities of individuals with ASD [19]. Following our research results, we have proposed a classification of the technologies used in ASD therapy based on their features and characteristics. It is presented in Fig. 7.

Hardware technologies include robots and electronic devices that have computerized physical components such as sensors and actuators. These devices can be programmed to provide real-time feedback, detect and respond to gestures and movements, or provide support based on individual patient needs. On the other hand, software technologies represent packages of software modules containing instructions, documents, and procedures that perform various tasks, these being desktop applications, mobile applications, web applications or augmented reality (AR), or virtual reality (VR) applications. Software technologies are typically used to support therapists and patients in the treatment process, allowing therapy to be customized according to patients' specific needs.

1) Desktop applications: Desktop applications used in autism therapy are software programs that run on personal computers or laptops, most commonly having Microsoft Windows, Apple Mac OS and Ubuntu Linux operating systems. Desktop applications are used to improve the communication, learning and development skills of individuals with autism. These applications include interactive games, educational applications and communication applications. Desktop applications are characterized by high performance, because they have enough memory to support complex user interfaces with numerous graphic elements and resource-consuming animations. The applications are easily adaptable to the individual needs of the users and the learning experience can be customized according to the skill level and preferences. The accessibility of technology is an important aspect in their frequent choice for ASD therapy, as there is no need to purchase expensive equipment.

2) Mobile applications: Mobile applications are an important part of autism therapy, giving users access to a variety of learning, communication and development tools. These applications are developed to be compatible with the most popular mobile operating systems, such as iOS or Android, and can be downloaded and installed on users' mobile devices: phones or tablets. The technological features of mobile applications for autism therapy include advanced tactile interaction such as multi-touch that allows the application to be used by several patients simultaneously, this versatility enables the use in both individual and group therapies.

In the Google Play and App Store, there is a large number of commercial applications dedicated to individuals with ASD. Most of them are educational games [17], visual schedule software [20], and tools for progress tracking [21]. Many academic researchers have also created functional prototypes targeting mobile devices. Common goals include support for development of language-communication skills, daily life skills, vocational-related skills, and social and emotional interaction [22]. A study conducted by Khaled Jedoui et al. from Stanford University School of Medicine [23] presents the mobile application "Guess What?". "Guess What?" is a mobile game available for Android and iOS platforms, designed to be a shared experience between the child, who tries to act out the request shown on the screen through gestures and facial expressions, and the parent, who has the task of guessing the word associated with the prompt within a game session time of 90 seconds. The study focuses on extracting emotion-tagged frames from video footage to train emotion classifiers that can adapt game difficulty and provide real-time feedback to the child. The study results show that the proposed tagging technique surpassed existing commercial emotion recognition APIs. The approach achieved an accuracy of 83.4% in labeling frames, a significant improvement over the best API's accuracy of 62.6%. The research presents a promising approach for using mobile games and automatic emotion labeling algorithms to provide social instruction to children with ASD.

3) Web applications: A web application is a software program that communicates via the World Wide Web and delivers web-based information to the user in HTML format [24]. The intrinsic features of web applications make them a complete solution for online and offline use. One of the main advantages of web technologies is their accessibility and for this reason, they are widely used in ASD therapy. They can be accessed from anywhere with an internet connection, making them particularly useful for individuals who live in rural or remote areas or who have limited access to logistic facilities. Therapists and educators prefer web applications for the educational content customization often provided by the platforms [25]. Educational content is tailored to the individual's learning abilities and progress. Visual aids such as images and videos are used in creating websites to engage and help people with ASD better understand concepts.

4) Augmented reality applications: Augmented reality technology uses a virtual environment that is overlaid on the real environment. This is done using video cameras These applications include interactive games, educational applications and communication applications. Desktop applications are characterized by high performance, because they have enough memory to support complex user interfaces with numerous graphic elements and resource-consuming animations. that are built into the devices. These cameras film the real environment and the AR software processes the image and adds virtual elements to it. Therefore, the user can see the real environment enhanced with virtual elements. The hardware used in AR technology includes video cameras, motion sensors, and specialized devices such as AR glasses, for instance Microsoft's HoloLens and Magic Leap One allow users to see virtual elements while moving through the real environment. AR technology is used in autism therapy to help autistic individuals develop social and communication skills. A common use case is to create scenarios for social interaction, such as engaging with virtual characters or participating in role-playing games. AR helps individual to understand and remember information, and it is not limited to one age group or level of education [26].

5) Virtual reality applications: Virtual reality is attracting increasing attention in the medical and healthcare industry, as it provides fully interactive three-dimensional simulations of real-world settings and social situations, which are particularly suitable for cognitive and performance training, including social and interaction skills [27]. Almost two decades ago, VR has been introduced as an effective tool in neurocognitive rehabilitation of patients with ASD [28]. VR applications include equipment such as a head-mounted display (HMD), haptic devices , hand controllers, foot controllers, non-body controllers, wireless trackers, and wrap-around displays. VR equipment allows the user to interact with the computer-generated environment in a seemingly real or physical way. Therapeutic applications of VR are based on the theory that the brain can process information more effectively when it is presented through a combination of sight, sound, and touch [29]. Owing to these distinctive features, therapists use VR to create scenarios that allow individuals affected by ASD to interact with virtual characters and learn to communicate. A study showed that the use of virtual reality technology in autism therapy can improve social and communication skills [30]. Academic researchers have shown interest in the impact VR technology can have on life and motor skills. In a study conducted by Tzanavari et al. [31], VR was employed to instruct six children with ASD on the safe method of crossing the street. The researchers observed that the children successfully acquired the necessary skills through the simulation and were able to apply them in real-world situations. VR technologies are a powerful tool for autism therapy. These technologies allow the creation of a safe and controllable environment for individuals with autism, where they can learn and develop their social and communication abilities.

6) Robots: Robot-assisted autism therapy (RAAT) is a method of therapy that uses specially designed robots to help individuals with autism spectrum disorders develop their social, communication, and cognitive skills. In RAAT, several types of robots are used: humanoid robots, animaloid robots, toy robots, and machine robots.

7) Humanoid robots: Humanoid robots are robots that resemble human beings, having hands, legs, and head with well-defined facial characteristics, including mouths and large eyes. These robots can be programmed to mimic human behavior and gestures, as well as to simulate human emotions. A research study conducted by Feng Wu et al. [32], aimed to replicate emotions by developing a humanoid robot equipped with thermoplastic elastomer (TPE) skin, which closely resembles human skin. The lifelike appearance of this robot had significant implications for clinical rehabilitation in the context of ASD. It provided a valuable platform for training individuals with autism to identify and mimic facial expressions, thus enhancing their emotional and social skills. Studies [33], [34] have also addressed this area of research and showed that individuals with ASD interpreted robots as a new intelligible species that present anthropomorphic thinking, that can have emotions and feel physical pain and treated them equitably as human beings. Humanoid robot can engage autistic people in ways that demonstrate essential aspects of human interaction, guiding them in therapeutic sessions to practice more complex forms of interaction found in social human-to-human interactions [35].

8) Animaloid robots: Animaloid robots are a category of robots that mimic the appearance of animals, such as dogs or cats, by using zoomorphic features such as fluffy fur, whiskers, a tail, or a beak. These robots are frequently used in therapy for children with autism spectrum disorders to improve their social and interaction skills. Because animal like robots have a playful appearance and often resemble pets, they are more attractive to children with autism, making them more receptive to their use in therapy [36]. Therapists use the animaloid robots to calm the child's emotional state and teach them how to interact with non-speaking beings, both from the perspective of empathy and to prevent exposure to danger. Through these robots, autistic children can learn how to behave with animals and learn how to develop their social skills [37], such as verbal and non-verbal communication, but also understand their emotions and needs [25].

9) Toy robots: Toy robots are a more affordable category of robots than other robots used in autism therapy, making them a popular choice for home use. The robotic toys come in a variety of sizes and shapes [35], often representing fantastic or imaginary characters, which makes them even more attractive to children. Toy robots help children to learn and perform tasks, thereby improving their cognitive skills [38]. These robots can also help develop motor skills and hand-eye coordination, as well as improve problem-solving skills. In addition, toy robots can be used for entertainment purposes [39], providing children and adults with a variety of fun activities.

10)Devices: In the context of autism therapy, devices are employed to provide assistance and help to improve the physical or mental capabilities of individuals with ASD, whether they are children or adults. They are classified into four main categories: tactile devices, speech devices, mobility devices, and head devices.

11)Tactile devices: Touch plays a crucial role in social communication and interactions and may be severely affected by the challenges in tactile perception, which are commonly observed in children with ASD. This results in either hyper- or hypo-sensitivity in these children [40]. Most tactile devices use vibrotactile sensors, pneumatic, and heat pump actuation [40]. Hardware capabilities of tactile devices enable complex simulations of different actions such as touching and hugging, and also provide the ability to capture fine finger touches. Tactile devices include toys with various textures and clothing with pressure sensors. The devices are frequently used in the therapy of people who have self-harming tendencies. They also help individuals with ASD with sensory stimulation and behavioral regulation. An example of smart clothing is "TellMe" created by Helen Koo from the University of California [41]. This clothing is specifically designed to address ASD and encourage boys who suffer from it to express themselves and enhance their communication skills. It incorporates therapeutic features such as sensors and actuators. The clothing includes a pressure sensor, a light sensor, and a motion sensor, along with actuators such as light-emitting diodes (LEDs), a direct current (DC) motor, and a vibration motor. Through interaction with interactive robot characters on the clothing, children can engage in activities like speaking into a microphone or triggering sensors and actuators. This interactive experience enables children to learn and practice expressing their feelings, emotions, and opinions. The clothing aims to provide a joyful and fascinating experience for children with ASD, stimulating self-confidence and self-expression .

12)Speech devices: Speech devices are technologies that help individuals with autism to communicate more effectively. These include speech synthesizers, tablets with specialized software [42], and hand-held pictographic devices for augmentative and alternative communication (AAC). Speech devices use advanced technology such as text to speech (TTS) and speech to text (STT) for natural language processing. These innovative devices prove to be invaluable resources for people who face language-related challenges, as they enable seamless communication with others, express emotions and interact with environment in a more meaningful way.

13)Mobility devices: Mobility devices are designed to assist individuals with ASD who have difficulty with mobility and motion. These devices range from simple walkers to advanced robotic exoskeletons. The hardware capabilities of mobility devices vary depending on the specific device. Some mobility devices are equipped with sensors and advanced software to help individuals with ASD improve their balance and coordination. These devices can also help individuals with ASD to become more independent and improve their overall quality of life.

14)Head devices: Head devices are used to address sensory and motor challenges faced by individuals with ASD. These devices include head-mounted displays and virtual reality headsets, which use advanced software and hardware to simulate real-world environments and activities. The hardware features of head devices include a variety of sensors that can monitor brain activity, stress levels and other aspects of psychological health. These devices also come with headphones that can play sounds. Head devices are specially designed to help autistic people focus, calm down and interact more easily with the environment through various techniques such as meditation and relaxing music.

Researchers' interest in the use of head-mounted devices increased in the past years due to the availability of affordable, consumer-grade HMD-based VR systems [43]. Dennis P. Wall et al. from Stanford University explored the power of head devices and developed a system named "SuperpowerGlass" [44]. This wearable aid is designed to assist children with ASD. The system uses Google Glass and an Android phone to provide children with real-time social cues. It includes various activities such as "Capture the Smile" and "Guess the Emotion" to engage children and promote emotional recognition and social interaction skills. The system underwent a 3-month study involving 14 families, and the results showed that children with ASD responded well to wearing the system at home and preferred the most expressive feedback option. The study also highlights an increased involvement of children in the wearable therapy sessions.

15)Interactive technologies in practice: In this study, we applied the unified theory of acceptance and use of technology (UTAUT) principles to understand why certain technologies are preferred over others and how they are used in practice. UTAUT is a model for user acceptance of information technology toward a unified view that explains user intentions pertaining to technology and subsequent usage behavior [12]. The theory states that there are four key constructs: (a) performance expectancy, (b) effort expectancy, (c) social influence, and (d) facilitating conditions, where the first three are direct determinants of usage intention and behavior, and the fourth is a direct determinant of user behavior [17].

The survey results, from cross-sectional study revealed that 55.9% of parents, therapist and teachers use technologies in ASD context. The remaining 44.1% of participants do not use technology, and one of the main reasons mentioned is the high price of devices, robots and applications.

The statistical analysis of participants' purchasing behavior revealed that 25% chose to purchase technology through direct payments, while the majority preferred to explore and use free technology. Further examination of the data indicates that participants' preferences for technology acquisition varied based on price ranges:

- 7% of participants preferred technology priced between €1 and €10;
- 10% of participants selected technology priced between €20 and €50;
- 3% of participants chosen technology priced between €50 and €100;
- 3% of participants purchased technology priced between €100 and €200;
- 2% of participants invested in technology priced above €400.

These findings are supported by Fig. 8, which visually represents the distribution of participants' preferences for technology acquisition.

A significant proportion of study participants, comprising over 40%, expressed a preference for mobile applications as their favored technology. This preference is primarily attributed to the ease of installation on personal smartphones or tablets, coupled with the convenience of accessing these applications from any location. In addition to mobile applications, speech devices, and desktop applications are commonly used by parents, caregivers and therapists in the context of therapy Fig. 9. These technologies are perceived to be effective in facilitating the development of communication skills and promoting social interaction of children with ASD.

In the survey, we conducted an inquiry into the specific purposes for which participants use interactive technologies, aiming to gain insights into the objectives associated with their usage. The findings indicate the following distribution of purposes among the participants:

- Educational purposes were reported by the majority, constituting 30% of the participants;
- Entertainment purposes accounted for 26% of the participants' responses;
- Communication purposes were reported by 17% of the participants;
- Therapeutic purposes were indicated by a relatively lower percentage, with only 15% of participants mentioning them;
- Diagnostic purposes were reported by merely 2% of the participants.
- Assistance purposes had the lowest percentage, with only 1% of participants indicating their usage for this objective.

Fig. 10 highlights the prominent utilization of technologies for educational and entertainment purposes.

Technologies used for educational purposes include learning platforms, video tutorials and applications that help children acquire new knowledge and skills. On the other hand, technologies used for entertainment purposes include video games, social media and music apps.

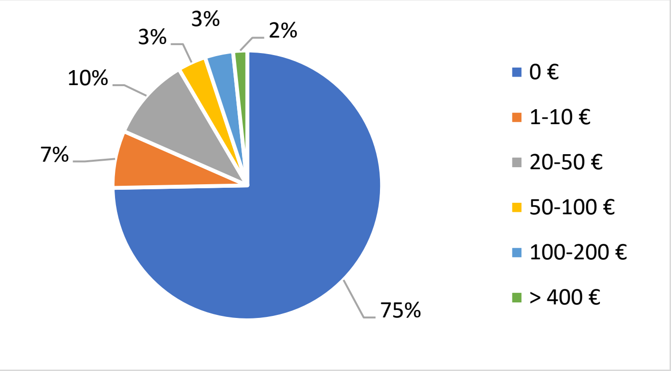


Fig. 8. The cost of technologies purchased by study participants.

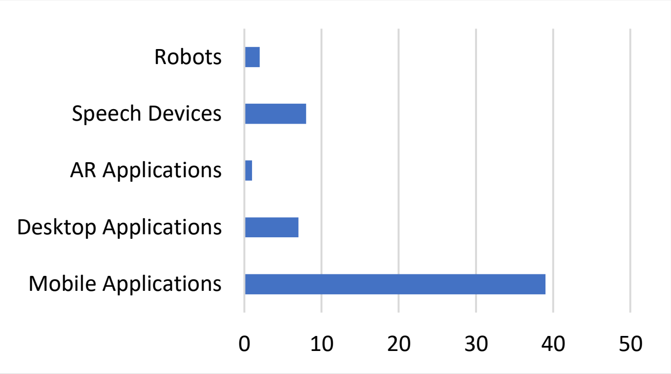


Fig. 9. Technologies practiced in real-life by parents, therapists and teachers.

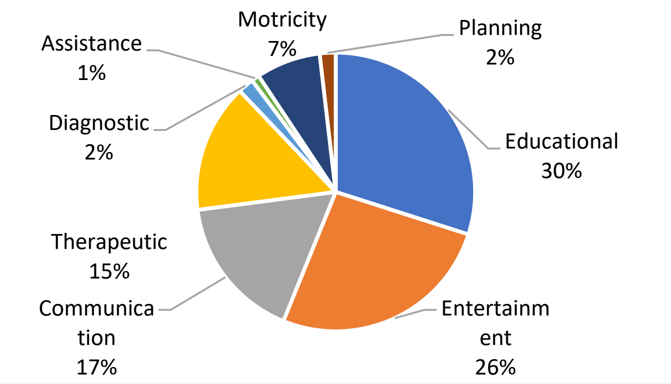


Fig. 10. The purposes for which the technologies were used by study participants.

Fig. 11. Interactive technologies used by study participants (Logos depicted in the graphic were sourced from official websites).

Fig. 11 presents some of the most frequently used interactive technologies by parents, therapists, and teachers to support the development of children with special needs. These technologies include apps like Leeloo AAC, AutiSpark, LetMeTalk, ABA Kit, and Autism ABC, which are specifically designed to help children with ASD communicate and improve their social and learning skills.

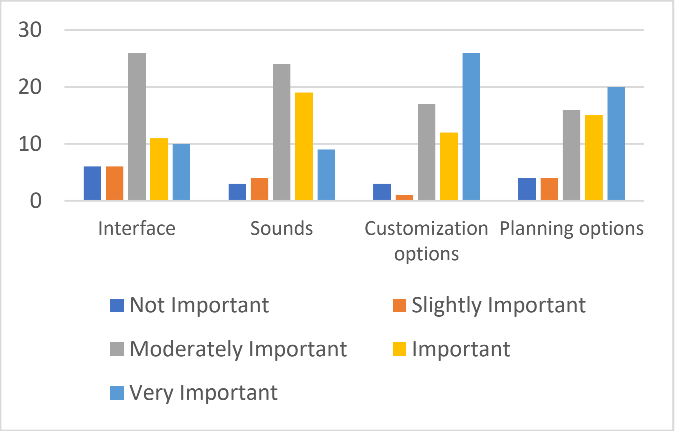
According to experts, interactive technologies are most frequently used in ABA therapy and in communication, speech and language therapy.

C. *Characteristics of Technologies in ASD Therapeutic Interventions*

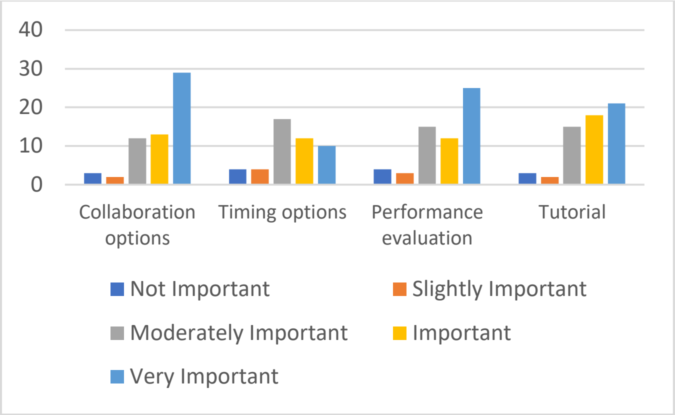
Improving skills of individuals with autism is a complex task that requires a personalized approach, taking into account the specific needs and characteristics of each individual.

The analysis of data from the cross-sectional study has highlighted the importance of different characteristics of interactive technologies used in ASD therapy.

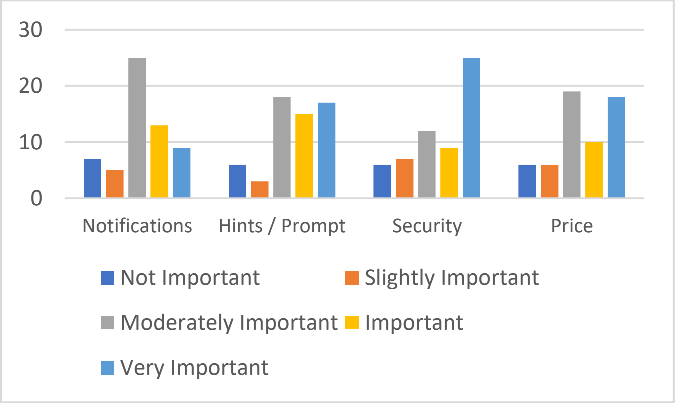
The participants assessed the importance of these technological characteristics, as depicted in Fig. 12. The data analysis unveiled that six characteristics were rated as "very important": collaboration options between parents and therapists; customization options for educational content; evaluation options for assessing the performance and progress of ASD patients; planning options; data security; and tutorial.



(a)



(b)



(c)

Fig. 12. Assessment of significance of technological characteristics. ASD experts' perspectives on (a) visual and audio design, customization of educational content and planning (b) collaboration and timing options, progress monitoring and evaluation, tutorial (c) notifications, hints, security, and price.

The characteristics that received evaluations of "moderately important" included: interface customization options; diagnostic options; timing options; notification; avatar; colorful UI interface; background music/sounds; price. The results did not indicate any prevailing evaluations for the characteristics evaluated as "slightly important" and "not important".

By combining the findings of the cross-sectional study and expert opinion, the most important characteristics of technologies used in ASD context were identified:

- *Visual and audio design*: Visual and audio design play a vital role in helping individuals with autism comprehend presented information. Interactive technologies with engaging design stimulate the senses and facilitate learning.
- *Planning and scheduling options*: Planning and scheduling options are important because individuals with autism may struggle with organizing and planning tasks. Interactive technologies provide planning options to assist ASD patients in managing their time efficiently and performing tasks more effectively.

- *Personalization of educational content*: Personalization of educational content is crucial because each individual with autism has specific educational needs and learns best through different methods. Interactive technologies can be customized for each individual, providing tailored educational content that meets their specific needs and abilities.

- *Progress monitoring and evaluation*: Progress monitoring and evaluation are essential to track the progress of individuals with autism in therapy. Interactive technologies provide monitoring and assessment tools that enable therapists and parents to follow the progress of individuals with autism.

- *Collaboration options*: Collaboration options with therapists, doctors, and parents are important because they must work together to provide appropriate therapy for individuals with autism. Interactive technologies can be designed to facilitate collaboration and communication between these stakeholders, thereby improving the quality of therapy provided to individuals with autism.

IV. Limitations

Our study focuses on interactive technologies that are designed for ASD therapy. Technologies can be a valuable tool in developing the aptitudes of patients with ASD, but their impact is not sufficiently investigated worldwide. Although we had a good representation of study participants, they were limited to parents, therapists, and teachers from a single county in Romania. The small number of subjects participating in the cross-sectional study, may affect the conclusions and their general applicability. To address this, future research will aim to include a larger and more diverse sample of participants to cover a wide spectrum of sociocultural experiences.

V. Conclusions

This research began by analyzing the ASD diagnosis and understanding that "spectrum" is a key term that characterizes the variety of forms and disabilities that individuals with ASD exhibit. Thus, we understood that the patient is the central element of our research, and technology is just a way to help the patient improve certain skills. Regardless of the degree of novelty and innovation of the technological solution, it must move from paper to practice and prove its effectiveness through well-documented results.

Individuals diagnosed with ASD show an increased interest in using technology for various purposes such as entertainment, communication, planning, or education. The positive attitude towards technology makes their use feasible in therapies by augmenting the information presented, simulating the visual, tactile and auditory senses, or monitoring the patient's physiological parameters. Experts in the field have highlighted that applied behavior analysis, communication, speech and language therapy, occupational therapy, and physical therapy are the therapies that most often involve technology in therapeutic procedures. The technology designed for use in the ASD context has various hardware specifications and software characteristics, depending on which they have been grouped into applications, robots, and devices. Current body of evidence supports the effectiveness of interactive technologies in improving ASD impairments such as communication deficits, social interaction deficits, motor disabilities, and mental retardation. Despite the obvious benefits, the participants in the study expressed concerns about the negative effects of technology use, such as agitation, aggression, and difficulty in use, associated with non-respect of the ISO standards in terms of appropriate visual or auditory design for persons with disabilities.

The multidisciplinary team that participated in this research consisted of therapists, teachers, software engineers, and parents of children diagnosed with ASD. Each participant brought their vast expertise to identify the important characteristics for ASD-specific technology. Thus, the adaptability of the technology to the patient's progress is by far the greatest need and challenge.

In response to cross-sectional study survey's question, which invited the participants to provide feedback and additional information not covered in the survey, a parent expressed a hopeful desire: "*I hope an application will be developed in Romania to assist parents, who don't have enough financial resources, to work with their children at home*.". This desire aligns perfectly with the goal of our future research, which is to leverage the valuable insights gained from the current study to create an accessible application that can support children with ASD and their families.

In conclusion, we encourage researchers to explore and develop innovative technological interventions for the ASD treatment, especially considering the growing prevalence of this neuropsychiatric condition.

Acknowledgment

This work was supported by European Regional Development Fund, Competitiveness Operational Programme 2014-2020, project "119722/ Centru pentru transferul de cunoştinţe către întreprinderi din domeniul ICT—CENTRIC, Contract subsidiar 22080/05.10.2022, Autism ASSISTant—Asistent virtual pentru dezvoltarea abilităților cognitive ale copiilor cu patologie de spectru autist", contract No. 5/AXA 1/1.2.3/G/13.06.2018, code SMIS 2014+ 119722 (ID P_40_305).

References

[1] L. Kanner, "Autistic disturbances of affective contact," Acta Paedopsychiatr, vol. 35, no. 4, pp. 100–136, 1968.

[2] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. American Psychiatric Association, 2013. doi: 10.1176/appi.books.9780890425596.

[3] World Health Organization, "Autism." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders

[4] P. Boucher, Assistive technologies for people with disabilities: in-depth analysis. Brussels: European Parliament, 2018.

[5] R. Sivakumar, "Google Forms In Education," Journal of Contemporary Educational Research and Innovations, vol. 9, pp. 35–39, 2019.

[6] L. Huang-Storms, Autism Spectrum Discorder Handbook.

[7] "Behavioral Management Therapy for Autism," NICHD Information Resource Center. [Online]. Available:

https://www.nichd.nih.gov/health/topics/autism/conditioninfo/treatments/behavioral-management

[8] R. Paul, "Interventions to Improve Communication in Autism," Child and Adolescent Psychiatric Clinics of North America, vol. 17, no. 4, pp. 835–856, Oct. 2008, doi: 10.1016/j.chc.2008.06.011.

[9] J. Case-Smith and M. Arbesman, "Evidence-Based Review of Interventions for Autism Used in or of Relevance to Occupational Therapy," The American Journal of Occupational Therapy, vol. 62, no. 4, pp. 416–429, Jul. 2008, doi: 10.5014/ajot.62.4.416.

[10] S. A. Cermak and A. E. Borreson, "Occupational Therapy," in Health Care for People with Intellectual and Developmental Disabilities across the Lifespan, I. L. Rubin, J. Merrick, D. E. Greydanus, and D. R. Patel, Eds., Cham: Springer International Publishing, 2016, pp. 1053–1067. doi: 10.1007/978-3-319-18096-0_90.

[11] S. Pegg, K. Hill, A. Argiros, B. O. Olatunji, and A. Kujawa, "Cognitive Behavioral Therapy for Anxiety Disorders in Youth: Efficacy, Moderators, and New Advances in Predicting Outcomes," Curr Psychiatry Rep, vol. 24, no. 12, pp. 853–859, Dec. 2022, doi: 10.1007/s11920-022-01384-7.

[12] Venkatesh, Morris, Davis, and Davis, "User Acceptance of Information Technology: Toward a Unified View," MIS Quarterly, vol. 27, no. 3, Art. no. 3, 2003, doi: 10.2307/30036540.

[13] N. Hass-Cohen and J. C. Findlay, Art therapy & the neuroscience of relationships, creativity, & resiliency: skills and practices, First edition. in The Norton series on interpersonal neurobiology. New York: W.W. Norton & Company, 2015.

[14] "Art Therapy for People on the Autism Spectrum," Disabled Living. [Online]. Available: https://www.disabledliving.co.uk/blog/art-therapy-for-people-on-the-autism-spectrum

[15] R. Van Kessel et al., "Autism and education—The role of Europeanisation in SOUTH-EASTERN Europe: Policy mapping in Bulgaria, Romania and Croatia," Children & Society, vol. 37, no. 5, pp. 1658–1671, Sep. 2023, doi: 10.1111/chso.12632.

[16] P. Michel, The Use of Technology in the Study, Diagnosis and Treatment of Autism. 2004.

[17] S. Alarcon-Licona and L. Loke, "Autistic Children's Use of Technology and Media: A Fieldwork Study," in Proceedings of the 2017 Conference on Interaction Design and Children, Stanford California USA: ACM, Jun. 2017, pp. 651–658. doi: 10.1145/3078072.3084338.

[18] S. Y. Kim et al., "A Systematic Quality Review of Technology-Aided Reading Interventions for Students With Autism Spectrum Disorder," Remedial and Special Education, vol. 43, no. 6, pp. 404–420, Dec. 2022, doi: 10.1177/07419325211063612.

[19] C. Wong et al., "Evidence-Based Practices for Children, Youth, and Young Adults with Autism Spectrum Disorder: A Comprehensive Review," J Autism Dev Disord, vol. 45, no. 7, pp. 1951–1966, Jul. 2015, doi: 10.1007/s10803-014-2351-z.

[20] J. Muchagata and A. Ferreira, "Visual Schedule: A Mobile Application for Autistic Children - Preliminary Study:," in Proceedings of the 21st International Conference on Enterprise Information Systems, Heraklion, Crete, Greece: SCITEPRESS - Science and Technology Publications, 2019, pp. 452–459. doi: 10.5220/0007732804520459.

[21] I. U. Rehman et al., "Features of Mobile Apps for People with Autism in a Post COVID-19 Scenario: Current Status and Recommendations for Apps Using AI," Diagnostics, vol. 11, no. 10, Art. no. 10, Oct. 2021, doi: 10.3390/diagnostics11101923.

[22] C. Putnam, C. Hanschke, J. Todd, J. Gemmell, and M. Kollia, "Interactive Technologies Designed for Children with Autism: Reports of Use and Desires from Parents, Teachers, and Therapists," ACM Trans. Access. Comput., vol. 12, no. 3, Art. no. 3, Sep. 2019, doi: 10.1145/3342285.

[23] H. Kalantarian, K. Jedoui, P. Washington, and D. P. Wall, "A Mobile Game for Automatic Emotion-Labeling of Images," IEEE Trans. Games, vol. 12, no. 2, Art. no. 2, Jun. 2020, doi: 10.1109/TG.2018.2877325.

[24] "GIS Dictionary," Contact Technical Support. [Online]. Available: https://support.esri.com/en-us/gis-dictionary/web-application

[25] M. L. da Silva, D. Gonçalves, T. Guerreiro, and H. Silva, "A Web-based Application to Address Individual Interests of Children with Autism Spectrum Disorders," Procedia Computer Science, vol. 14, pp. 20–27, 2012, doi: 10.1016/j.procs.2012.10.003.

[26] M. Wedyan et al., "Augmented Reality for Autistic Children to Enhance Their Understanding of Facial Expressions," MTI, vol. 5, no. 8, Art. no. 8, Aug. 2021, doi: 10.3390/mti5080048.

[27] M. Zhang, H. Ding, M. Naumceska, and Y. Zhang, "Virtual Reality Technology as an Educational and Intervention Tool for Children with Autism Spectrum Disorder: Current Perspectives and Future Directions," Behavioral Sciences, vol. 12, no. 5, Art. no. 5, May 2022, doi: 10.3390/bs12050138.

[28] C. G. Trepagnier, "Virtual environments for the investigation and rehabilitation of cognitive and perceptual impairments," NRE, vol. 12, no. 1, Art. no. 1, Feb. 1999, doi: 10.3233/NRE-1999-12107.

[29] V. Pandey and L. Vaughn, "The Potential of Virtual Reality in Social Skills Training for Autism: Bridging the Gap Between Research and Adoption of Virtual Reality in Occupational Therapy Practice," The Open Journal of Occupational Therapy, vol. 9, no. 3, Art. no. 3, Jul. 2021, doi: 10.15453/2168-6408.1808.

[30] S. Parsons and P. Mitchell, "The potential of virtual reality in social skills training for people with autistic spectrum disorders: Autism, social skills and virtual reality," Journal of Intellectual Disability Research, vol. 46, no. 5, Art. no. 5, May 2002, doi: 10.1046/j.1365-2788.2002.00425.x.

[31] A. Tzanavari, N. Charalambous-Darden, K. Herakleous, and C. Poullis, "Effectiveness of an Immersive Virtual Environment (CAVE) for Teaching Pedestrian Crossing to Children with PDD-NOS," in 2015 IEEE 15th International Conference on Advanced Learning Technologies, Hualien, Taiwan: IEEE, Jul. 2015, pp. 423–427. doi: 10.1109/ICALT.2015.85.

[32] F. Wu, S. Lin, X. Cao, H. Zhong, and J. Zhang, "Head Design and Optimization of An Emotionally Interactive Robot for the Treatment of Autism," in Proceedings of the 2019 4th International Conference on Automation, Control and Robotics Engineering, Shenzhen China: ACM, Jul. 2019, pp. 1–10. doi: 10.1145/3351917.3351992.

[33] P. H. Kahn, H. E. Gary, and S. Shen, "Children's Social Relationships With Current and Near-Future Robots," Child Dev Perspect, vol. 7, no. 1, Art. no. 1, Mar. 2013, doi: 10.1111/cdep.12011.

[34] B. Szymona et al., "Robot-Assisted Autism Therapy (RAAT). Criteria and Types of Experiments Using Anthropomorphic and Zoomorphic Robots. Review of the Research," Sensors, vol. 21, no. 11, Art. no. 11, May 2021, doi: 10.3390/s21113720.

[35] A. Alabdulkareem, N. Alhakbani, and A. Al-Nafjan, "A Systematic Review of Research on Robot-Assisted Therapy for Children with Autism," Sensors, vol. 22, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/s22030944.

[36] J. Bharatharaj, L. Huang, R. Mohan, A. Al-Jumaily, and C. Krägeloh, "Robot-Assisted Therapy for Learning and Social Interaction of Children with Autism Spectrum Disorder," Robotics, vol. 6, no. 1, p. 4, Mar. 2017, doi: 10.3390/robotics6010004.

[37] C. M. Stanton, P. H. Kahn Jr., R. L. Severson, J. H. Ruckert, and B. T. Gill, "Robotic animals might aid in the social development of children with autism," in Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, Amsterdam The Netherlands: ACM, Mar. 2008, pp. 271–278. doi: 10.1145/1349822.1349858.

[38] M. H. Laurie, P. Warreyn, B. V. Uriarte, C. Boonen, and S. Fletcher-Watson, "An International Survey of Parental Attitudes to Technology Use by Their Autistic Children at Home," J Autism Dev Disord, vol. 49, no. 4, Art. no. 4, Apr. 2019, doi: 10.1007/s10803-018-3798-0.

[39] H. Kozima, M. P. Michalowski, and C. Nakagawa, "Keepon: A Playful Robot for Research, Therapy, and Entertainment," Int J of Soc Robotics, vol. 1, no. 1, pp. 3–18, Jan. 2009, doi: 10.1007/s12369-008-0009-8.

[40] J.-J. Cabibihan, H. Javed, M. Aldosari, T. Frazier, and H. Elbashir, "Sensing Technologies for Autism Spectrum Disorder Screening and Intervention," Sensors, vol. 17, no. 12, Art. no. 12, Dec. 2016, doi: 10.3390/s17010046.

[41] H. Koo, "'TellMe': therapeutic clothing for children with autism spectrum disorder (ASD) in daily life," in Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program,

Seattle Washington: ACM, Sep. 2014, pp. 55–58. doi: 10.1145/2641248.2641278.

[42] J. P. Solis and I. C. Valenzuela, “Speech Assistive Device For Students With Autism Spectrum Disorder: A Review,” Journal of Computational Innovations and Engineering Applications, vol. 4, no. 2, Art. no. 2, 2020.

[43] M. Schmidt, N. Newbutt, C. Schmidt, and N. Glaser, “A Process-Model for Minimizing Adverse Effects when Using Head Mounted Display-Based Virtual Reality for Individuals with Autism,” Front. Virtual Real., vol. 2, p. 611740, Mar. 2021, doi: 10.3389/frvir.2021.611740.

[44] P. Washington et al., “SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism,” Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 1, no. 3, Art. no. 3, Sep. 2017, doi: 10.1145/3130977.

Drier Bed Adsorption Predictive Model with Enhancement of Long Short-Term Memory and Particle Swarm Optimization

MarinaYusoff[1], Mohamad Taufik Mohd Sallehud-din[2], Nooritawati Md. Tahir[3], Wan Fairos Wan Yaacob[4], Nur Niswah Naslina Azid @ Maarof[5], Jasni Mohamad Zain[6], Putri Azmira R Azmi[7], Calvin Karunakumar[8]

Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor Darul Ehsan[1, 3, 4, 6, 7]

PETRONAS Research Sdn Bhd, Jln Ayer Hitam, Kawasan Institusi Bangi, 43000 Bandar Baru Bangi, Selangor[2]

College of Computing, Informatics and Media, Universiti Teknologi MARA Cawangan Kelantan, Kampus Kota Bharu, Lembah Sireh, Kota Bharu, 15050, Kelantan[5]

PETRONAS Research Sdn Bhd, Jln Ayer Hitam, Kawasan Institusi Bangi, 43000 Bandar Baru Bangi, Selangor[8]

***Abstract*—The drier bed adsorption processes remove moisture from gases and liquids by ensuring product quality, extending equipment lifespan, and enhancing safety in various applications. The longevity of adsorption beds is quantified by net loading capacity values that directly impact the effectiveness of the moisture removal process. Predictive modeling has emerged as a valuable tool to enhance drier bed adsorption systems. Despite the increasing significance of predictive modeling in enhancing the efficiency of drier bed adsorption processes, the existing methodologies frequently exhibit deficiencies in accuracy and flexibility, which are crucial for optimizing process performance. This research investigates the effectiveness of a hybrid approach combining Long Short-Term Memory and Particle Swarm Optimization (LSTM+PSO) as a proposed method to predict the net loading capacity of a drier bed. The train-test split ratios and rolling origin technique are explored to assess model performance. The findings reveal that LSTM+PSO with a 70:30 train-test split ratio outperform other methods with the lowest error. Bed 1 exhibits an RMSE of 1.31 and an MSE of 0.91, while Bed 2 archives RMSE and MSE values of 0.81 and 0.72, respectively and Bed 3 with an RMSE of 0.19 and an MSE of 0.13, followed by Bed 4 with an RMSE of 0.67 and an MSE of 0.36. Bed 5 exhibits an RMSE of 0.42 and an MSE of 0.34. Furthermore, this research compares LSTM+PSO with LSTM and conventional predictive methods: Support Vector Regression, Seasonal Autoregressive Integrated Moving Average with Exogenous Variables, and Random Forest.**

Keywords*—Adsorption; Long Short-Term Memory; net loading capacity; Particle Swarm Optimization; prediction

I. Introduction

Drying is used in various sectors, including agriculture, pharmaceuticals, energy, and other process industries, to transform liquid within a product into a solid [1], [2]. External energy, such as fossil fuel, generates high temperatures in a drying system's rotary bet, fluidized bed, spray, tray, impinging jet, and pulse combustion dryers [2]–[4]. This method involves drying wet goods at high temperatures [3], [4] to ensure the best possible performance. Drying techniques range from the sun to the oven, freezer, and microwave [3], [5]. Adsorption shows more promise than other drying methods when it comes to air drying [6].

The drier bed adsorption process is widely used across various industries [7], [8]. It is an essential parameter for estimating the remaining life of the drier bed and must coincide with the planned plant shutdown period. The overestimation of bed capacity will lead to unexpected moisture breakthroughs, resulting in production losses due to the drying out of equipment further down the line. In this research, Net Loading Capacity (NLC) measures the amount of moisture an adsorption bed can absorb before replacing or regenerating. An adsorption bed with a longer lifespan can extract more water before needing to be replaced or regenerated, which can maximize efficiency and save expenses. The challenge of accurately measuring dryer process system performance remains unaddressed despite its critical role in guaranteeing energy conservation, process dependability, and product quality [2, 4, and 9].

In recent years, Long Short-Term Memory (LSTM) neural networks are also known for their ability to model sequential data effectively, such as time series or text data. LSTM also has advanced potential for prediction modelling in various domains such as prediction of crude oil [10]–[12], financial [13], energy consumption [14]–[16], and medical [17]. Their capability to capture long-range dependencies and adapt to changing input patterns makes them a promising tool for modelling and predicting the dynamic behaviour of drier bed adsorption. However, performing well requires a large number of data and substantial computational resources. This makes them less suitable for small datasets due to overfitting. Furthermore, it can be sensitive to hyperparameter settings since finding the appropriate parameter can be time-consuming.

The optimization techniques are crucial in improving the prediction efficiency of the models. Particle Swarm Optimization (PSO) is a robust optimization algorithm inspired by birds flying for food, which has been widely applied to various optimization problems [18]. The combination and integration of PSO with LSTM aims to tackle the information of particles to fine-tune the model parameters and optimize the prediction accuracy of the drier bed adsorption model of a small number of datasets.

The contributions of this paper are in the following:

1) A new particle representation is used for PSO implementation in an NLC drier bed performance-based molecular sieve.

2) A novel idea of combining LSTM with a new particle representation strategy of a PSO for predicting drier bed adsorption capacity, namely LSTM+PSO.

3) LSTM+PSO facilitates the analysis of temporal predictions, optimizing prediction models and improving prediction accuracy using a small number of adsorption bed life monitoring datasets.

4) A comparison analysis of LSTM+PSO with LSTM and traditional machine learning and statistical methods was performed.

The remainder of paper is organized by sections. Section II presents the related work. Section III delves into methodology detailing the architecture of the LSTM+PSO model and the optimization process. Section IV discusses the proposed solution, the model of LSTM+PSO performance with LSTM, SVR, and SARIMAX. Section V discusses the performance evaluation. Section VI gives the computational results. Section VII and Section VIII presents the discussion and conclusion respectively.

II. Related Work

Adsorption drying reduces the amount of water vapor in humid air by passing it through the solid adsorbent level of dehydration, which influences the selection of an appropriate adsorbent [6], [19]. Adsorbent bed dryers come in various configurations, including packed beds, coated channels, and annular coated tubes [20]. The bed's efficiency is affected by design parameters such as bed length and adsorbent mass [7]. The design of the drops must allow for high transfer rates while also being appropriately sized accommodating the allowable pressure drop [21]. Furthermore, the dryer should be stable over long periods of operation, have low toxicity, be corrosion resistant, and be cost-effective. Molecular sieves are an example of an adsorbent commonly used in natural gas plants [19]. The ability of water molecules to diffuse into the pores of the adsorbent limits the overall adsorption rate [6].

The advantages of adsorption drying include a low impact of temperature and pressure on the adsorption process, the use of simple equipment, reduced spatial demands, efficient humidification capabilities, the ability to use various heat sources for adsorbent regeneration, and cost-effective operation [6], [21]. The short adsorption-desorption cycles and low-temperature regeneration techniques are employed that can help achieve low operational costs [22]. Adsorption drying has several disadvantages, including increased heating costs, energy-intensive regeneration processes, and potential sorbent abrasion [9], [21]. Various approaches have been proposed to improve the limitations of estimating drying capacity to achieve optimal drying performance. These include the use of mathematical models, simulations of the drying apparatus, and machine learning techniques.

The use of a mathematical model for the drier indicates that it causes an increase in drying capacity to optimize the energy consumption of an amount of heat [23]. Furthermore, both the dryer efficiency and sustainability index demonstrate a significant reliance on the extent of heat recovery in the case of a spray dryer [24]. On the other hand, the significance of materials' drying behavior highlights limitations to enhance accuracy and reliability in adopting capillary active insulation materials [25]. The mathematical modeling for a multistage phosphate pellet roasting process offers the potential for significant energy savings [26]. However, there still needs to be more research on adsorption dryer beds, especially in the prediction of the NLC.

III. Material and Methods

This section explains the steps to address the predictive analytics challenge in drier bed adsorption, specifically in predicting NLC. The research framework begins with data acquisition, which involves data collection from five different beds: Bed 1, Bed 2, Bed 3, Bed 4, and Bed 5. Next, the data undergoes a pre-processing stage where unnecessary or irrelevant information, such as bad input and intf shut are eliminated to ensure data quality. Additionally, noisy data is filtered out to enhance the reliability of the dataset. The next step is featuring selection guided by correlation coefficients [27], which aims to identify the most relevant features for the predictive modeling task. The dataset is divided into training and testing subsets, setting the model development stage.

In the research model development phase, various methodologies comparing conventional methods such as LSTM, SVR, SARIMAX, and Random Forest are compared with the proposed LSTM+PSO. Fig. 1 provides a research framework. The data acquisition process involves the implementation of an automated procedure. In this research, computational analysis has been conducted focuses on the process within the drier beds for Bed 1, Bed 2, Bed 3, Bed 4, and Bed 5. The components that involve automatic breakthrough time identification, the identification and checking of the regeneration efficiency, and the identification regeneration cycle are all components of this step for each drier bed.

A. Data Pre-processing

Data pre-processing involves data extraction. The process of filtering historical data is handled by data extraction. A few steps in data processing are removing unnecessary data such as bad input, intf shut, and removing noisy data. Removing unnecessary data involves identifying and eliminating columns or variables within the dataset that either do not contribute valuable information or contain redundant information. During data processing, noisy data points are identified and filtered out.

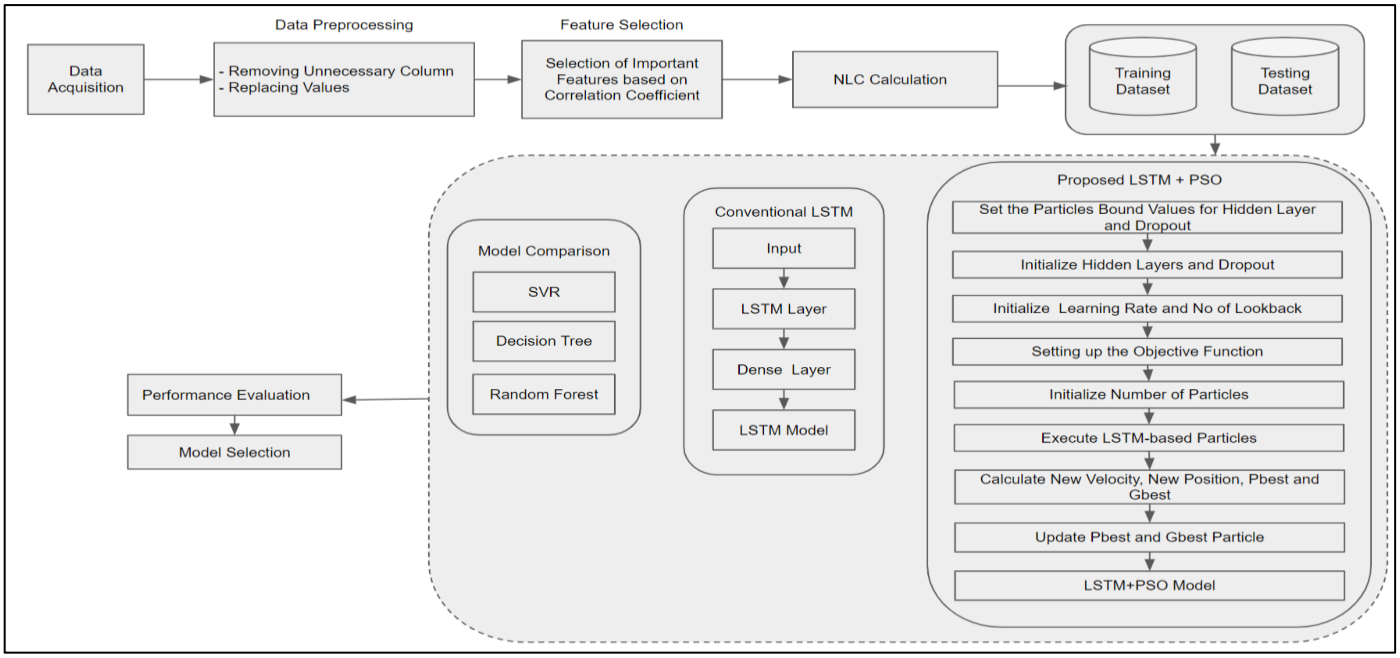

Data Preprocessing
Feature Selection
Data Acquisition
- Removing Unnecessary Column
- Replacing Values
Selection of Important Features based on Correlation Coefficient
NLC Calculation
Training Dataset
Testing Dataset
Proposed LSTM + PSO
Set the Particles Bound Values for Hidden Layer and Dropout
Initialize Hidden Layers and Dropout
Initialize Learning Rate and No of Lookback
Setting up the Objective Function
Initialize Number of Particles
Execute LSTM-based Particles
Calculate New Velocity, New Position, Pbest and Gbest
Update Pbest and Gbest Particle
LSTM+PSO Model
Conventional LSTM
Input
LSTM Layer
Dense Layer
LSTM Model
Model Comparison
SVR
Decision Tree
Random Forest
Performance Evaluation
Model Selection

Fig. 1. Research framework.

B. Features Selection

The features are chosen based on their correlation with beds 1, 2, and 3 to identify the most relevant ones. The correlation can use two sets of correlation values: {0.1, 0.2, 0.3, 0.4, 0.5} and {-0.1, -0.2, -0.3, -0.4, -0.5}, calculated using the Pearson correlation coefficient. Eq. (1) presents the Pearson formulas for the correlation coefficient [27]. This coefficient ranges from -1 to 1. The coefficient value equal to -1 indicates a perfect negative correlation. Meanwhile, 1 indicates a perfect positive correlation. In addition, a zero value indicates no linear correlation between the variables.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (1)$$

where, in these equations, r represents the correlation coefficient, x_i signifies the values of the variable of x in a sample, x denotes the mean (average) of the x variable values, y_i stands for the values of the y variable in a sample, and y represents the mean (average) of the y variable values.

C. Establishment of NLC Data

NLC data is crucial for predicting drier bed adsorption capacity. The values of NLC are drier bed. Historical datasets on drier bed adsorption were employed to build predictive and validation models. The calculation of NLC includes historical data features, regeneration processes, and breakthrough times for adsorption. Following the processes, the final datasets for each bed are less than 30. The challenge is to achieve good prediction accuracy with a small dataset.

IV. Proposed Solution

The proposed solution combines the strengths of LSTM and PSO for a more robust and practical approach for a small dataset. The LSTM network provides powerful sequence modeling capabilities. Meanwhile, PSO is used to fine-tune hyperparameters and optimize the performance of the LSTM model. This hybrid approach aims to utilize the strengths of both techniques to improve predictive accuracy and overall performance. Following are the details of the hybridization of LSTM+PSO.

A. Particle Representation

A real-value PSO implementation followed the initial PSO steps. This representation is new for PSO implementation, which supports a new strategy to be embedded in LSTM and a drier bed performance-based molecular sieve. Fig. 2 is the representation for the particle, $H = \{H_1, H_2, H_3, H_4, H_5 \ldots H_n\}$ and Dropout (DP). This representation defines the characteristics; include a number of layers including DP structure, within the optimization process in the search space.

H_1	H_2	DP

Fig. 2. Particle representation.

where, the parameters are defined as H_1 represents hidden layer 1 with a range of [0.80, 1.28], H_2 represents hidden layer 2 with a range of [0.80, 2.56], and DP with a range of [0.05, 0.50]. These parameter definitions specify the allowable values or ranges for each parameter within the optimization process of LSTM+PSO. Each particle component consists of layers and DP is randomly initiated in a population. A real value for all layers is initiated as in Eq. (2) and Eq. (3).

$$d = 2 * (np.random.rand(n, 1)) - 1 \quad (2)$$

$$d = rand(x, y) * z \quad (3)$$

Where, the variables are defined as x, y, and z represents indices used in layer calculations. x is a variable that takes values from the set $\{1, 2, 3, ..., m\}$. These indices are used for referencing through layers within a calculation. A particle update procedure was introduced, which is iteratively adjusted to identify the optimal model fit for the network. The dynamic range update is based on +8 or -8 for H_1 and H_2 values, and DP is an increased value of 0.01. Eq. (3) and Eq. (4) present the velocity and position formulas for the PSO, respectively.

$$V_{id\,(new)} = W * V_{id} + C_1 r_1 * \left(P_{best(id)}\right) - X_{id} + C_2 r_2 * \left(G_{best(id)} - X_{id}\right) \tag{4}$$

$$X_{id(new)} = X_{id} + V_{id\,(new)} \tag{5}$$

where, $V_{id(new)}$ represents the updated velocity, V_{id} stands for the current velocity, X_{id} denotes the current position, and $X_{id(new)}$ signifies the new position of a particle in a PSO. The parameters W represent the inertia weight, while C_1 and C_2 are the acceleration coefficients controlling the impact of personal and global best positions on particle movement. r_1 *and* r_2 are random functions or values used for stochastic behaviour. $P_{best(id)}$ represents the personal best position for a particle with the identifier id, and $G_{best(id)}$ signifies the global best position for the same particle identifier within the PSO algorithm.

B. LSTM+PSO Algorithm

A novel approach is a hybrid model that integrates the PSO algorithm and the LSTM; this model is intended to serve the oil and gas industry. Algorithm 1 provides an overview of the LSTM+PSO method. The LSTM+PSO optimize LSTM hyperparameters for prediction. The algorithm begins by initialization of particle values in Steps 1-2. This is vital for the PSO optimization process. Next, Steps 3-4 are initializing the particles with random values. Determine the number of past future data points in the LSTM model for Step 5.

Furthermore, Step 6 defines the objective function the PSO algorithm will optimize. This function measures the performance of the LSTM model on the given task. Initialize particles and iterations number in Steps 7-8. These steps specify the number of particles in the PSO optimization and set the maximum number of iterations for the PSO algorithm. Step 9 is loading the dataset that contains the selected features relevant to the task from the selected features. Step 10 is performing the feature scaling on the dataset to ensure that the data is within a consistent range, which is typically important for LSTM models. Split the dataset into training and testing sets for model evaluation and train the LSTM neural network with the initialized particles, which represent different configurations of the LSTM model in Steps 11-12.

Algorithm 1: LSTM+PSO

```
1   Begin
2   Initialize particles
3   Initialize hidden layers and dropout-based particle values
4   Initialize learning rate
5   Set the number of lookbacks, number of future points
6   Setting up the objective function
7   Initialize particles number
8   Initialize iterations number
9   Load dataset from the selected features
10  Features scaling for LSTM fitting
11  Set the Train Test Split
12  Execute LSTM
13  Calculate Pbest and Gbest values for each particle
14  Do
15    For each particle
16    Calculate the new velocity value, $V_{(new)}$
17    Calculate new position, $D_{(new)}$
18    Calculate Pbest $_{(new)}$
19    Calculate Gbest $_{(new)}$
20    For each particle dimension
21    If current Pbest > current Gbest
      **Update new particle**

22    If current Pbest < current Gbest
      **Update new particle**
23  While (stopping condition is reached)
24  End
```

Moreover, evaluate the performance of each particle (LSTM configuration) using the objective function and determine the personal best (P_{best}) and global best (G_{best}). Step 14-Step 23 begins a loop. Iterate through each particle. The particle's velocity is updated based on its current Pbest and Gbest positions. The position of the particle is updated using the new velocity. Re-evaluate the performance of the particle with its new position and update Pbest if it improves in Step 18. Step 29 is to update the global best (G_{best}) if a particle of Pbest is better than the current Gbest. Iterate through each dimension of the particle by adjusting the particle dimension by subtracting and adding a particle change value in Step 20-Step 22. Continue the loop until a stopping condition is met, such as reaching a maximum number of iterations in Step 23. It marks the endpoint when the stopping condition is met or when the maximum number of iterations is reached.

In this research, LSTM is further enhanced by embedding the advantages of PSO. In this study, LSTM and PSO are combined. PSO can assist in finding an optimal solution, such as in obtaining a better architecture of LSTM. Fig. 1 demonstrates LSTM + PSO architecture. The main steps are similar to LSTM, which defines the input size, hidden layer, and output size. Input size corresponds to the number of input sequences or several features. The hidden layer size specifies the number of hidden layers, and the output size is set to 1, which indicates the number of items in the output predicts the NLC. The PSO elements are embedded for LSTM architecture determination.

V. Performance Evaluation

The next step is to train the data and define the epochs number using the train-test split and rolling origin. RMSE measures the average error between predicted and actual values, with lower values indicating better model accuracy. MAE represents the average absolute error, whereas lower values also suggest better model accuracy. Three types of splitting are used: 70:30, 80:20, and 90:10, dividing between training and testing data. The model's performance will be evaluated by analyzing its ability to predict NLC.

The RMSE acceptance criteria [28]–[30] are categorized as follows: RMSE values falling within the range of ≤ 0.75 are considered very Good, while those between 0.75 and 1.0 are deemed Good. RMSE values ranging from 1.0 to 2.0 are labeled as satisfactory, and any RMSE exceeding 2.0 is categorized as unsatisfactory. These criteria provide a standardized assessment for evaluating the accuracy and quality of RMSE values in various applications or studies. RMSE equation is a mathematical expression used to quantify the average deviation between predicted values ($\hat{y}_i$) and actual values (y_i) within a regression model, as shown in Eq. (6).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{6}$$

The LSTM model also has the best performance based on the results of evaluation metrics MAE. The MAE acceptance criteria [31], [32], as referenced, are categorized as follows: MAE values between 0 and 3 are considered very good. Those falling within the range of 3 to 6 are labelled as good. MAE values between 6 and 9 are categorized as average, and if the MAE falls within the range of 9 to 12, it is referred to as variable data. Any MAE exceeding 12 is designated as higher variability. These criteria offer a standardized framework for assessing the quality and suitability of MAE values in different contexts or studies. The MAE equation is a mathematical formula used to quantify the variance between the prediction (y_i) and the real value (x_i) by dividing this variance by the square root of the number of data points in the observations (n), as shown in Eq. (7).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{7}$$

VI. Computational Results

The computational results involve using root-mean-squared error (RMSE) and mean-squared error (MAE) metrics. These metrics are applied and compared to conventional methods using the setting of parameters to assess the effectiveness and suitability of the algorithm.

A. Parameter Setting

Table I outlines the parameter settings for the LSTM+PSO model. It specifies the values assigned to different parameters for training and evaluating the model. The train-test split parameter determines how the dataset is divided for training and testing with options of 90% training and 10% testing, 80% training and 20% testing, or 70% training and 30% testing.

The rolling origin parameter indicates the number of time steps considered when making predictions with 1, 2, or 3 options. The epoch parameter sets the number of times it is passed forward and backward through the LSTM network during training, which is set at 30. The learning rate parameter defines the step size at which the model adjusts its weights during optimization, set to 0.1. Finally, the batch size parameter determines the number of data points used in each iteration of the training process and is set to 256.

TABLE I. Parameter Setting of LSTM+PSO

Parameter	Value
Train-Test Split	90:10, 80:20, 70:30
Rolling Origin	1,2,3
Epoch	30
Learning Rate	0.1
Batch Size	256
Population	30

B. Computational Results using LSTM+PSO

The comparison computational results achieved through the LSTM+PSO approach are presented in Table II. The data was split using a 90:10, 80:20, and 70:30 ratio, which utilized a lookback of three for Bed 1, a lookback of two for Bed 2, and a lookback of four for Bed 3 for prediction. The LSTM+PSO that utilize a Train-Test Split ratio of 70:30 has demonstrated its effectiveness across Bed 1, Bed 2, Bed 3, Bed 4, and Bed 5, as indicated by both RMSE and MAE evaluations.

TABLE II. Computational Result using Train-Test Split

Beds	Train-Test Split	RMSE	MAE
Bed 1	90:10	1.68	1.39
	80:20	1.53	1.16
	70:30	**1.31**	**0.91**
Bed 2	90:10	1.07	1.00
	80:20	0.97	0.90
	70:30	**0.81**	**0.72**
Bed 3	90:10	0.28	0.21
	80:20	0.21	0.15
	70:30	**0.19**	**0.13**
Bed 4	90:10	0.71	0.54
	80:20	0.70	0.56
	70:30	**0.67**	**0.36**
Bed 5	90:10	0.46	0.38
	80:20	0.45	0.37
	70:30	**0.42**	**0.34**

For Bed 1, an RMSE score of 1.31 falls within the satisfactory range, indicating that the predictive model provides reasonably accurate results for this bed. Furthermore, the MAE value of 0.91 is categorized as very good, signifying that the model's prediction closely aligns with the actual values. It demonstrates a high level of accuracy as compared to other percentages of Train-Test Split. Moving on to Bed 2, the RMSE score of 0.81 is good, implying that the model delivers accurate predictions with a relatively low margin of error.

Similarly, the MAE value of 0.72 is labeled as very good, indicating that the model's performance is highly accurate for Bed 2. In addition, for Bed 3, an RMSE score of 0.19 is considered satisfactory. This suggests that the accuracy of the model is acceptable. It is worth noting that the MAE value of 0.13 is once again classified as very good by underscoring the model's capability to make highly accurate predictions for Bed 3. Furthermore, in the case of Bed 4, the RMSE score of 0.67 is categorized as good. This indicates that the predictive model provides accurate predictions with a relatively small margin of error for Bed 4. Additionally, the MAE value of 0.36 is labeled as very good and highlights that the model's performance is highly accurate when predicting outcomes for Bed 4. Lastly, for Bed 5, the RMSE shows a good value, which is 0.42, along with the MAE obtained of about 0.34, which indicates good performance to the predictive model.

Table III presents an analysis of performance metrics, specifically RMSE and MAE, for various bed types (Bed 1, Bed 2, Bed 3) across rolling origin values of 1, 2, and 3. The LSTM+PSO approach applied rolling origin has consistently proven its effectiveness in predicting outcomes for Bed 1 until bed 3, as evidenced by the RMSE and MAE assessments. Notably, for Bed 1, the RMSE of 1.78 falls within the satisfactory range, while the MAE of 1.62 is categorized as very good. Similarly, for Bed 2, the RMSE of 0.82 is deemed suitable, and the MAE of 1.08 is labeled very good. For Bed 3, the RMSE of 0.70 is considered satisfactory, and the MAE of 0.14 is classified as very good.

TABLE III. COMPUTATIONAL RESULT USING ROLLING ORIGINS

Types of Beds	Rolling Origin	RMSE	MAE
Bed 1	**1**	**1.78**	**1.62**
	2	2.00	1.79
	3	2.46	2.19
Bed 2	**1**	**0.82**	**1.08**
	2	0.93	1.18
	3	1.17	1.28
Bed 3	**1**	**0.70**	**0.14**
	2	1.21	0.80
	3	1.42	0.38

C. *Computational Results with LSTM and Conventional Methods*

The comparison experimental results of conventional methods of LSTM, SVR, SARIMAX, Random Forest (RF), and LSTM+PSO are shown in Table IV. The choice of parameter settings for the split percentage of 70:30 is similar to the LSTM+PSO approach. A lower RMSE and MAE suggest superior predictive performance and indicate that the model's prediction is closer and aligned to the actual values.

In the case of Bed 1, the SARIMAX and RF models showed better results than the LSTM, SVR, and LSTM+PSO models. It was evident through their lower values of RMSE and MAE. Moreover, it can be observed that LSTM+PSO exhibits the highest predictive accuracy in terms of both RMSE and MAE metrics for Bed 2, Bed 3, Bed 4, and Bed 5. Regardless of SARIMAX and RF demonstrating lower values of RMSE and MAE, they are not suitable for forecasting due to their inability to accommodate the lookback value, resulting in unviable forecast points.

TABLE IV. COMPUTATIONAL RESULT WITH CONVENTIONAL METHOD

Types of Beds	Metrics	LSTM	SVR	SARIMAX	RF	LSTM+PSO
Bed 1	RMSE	1.45	2.81	1.18	1.10	**1.31**
	MAE	1.33	2.79	0.76	0.99	**0.91**
Bed 2	RMSE	0.69	1.08	4.68	1.12	**0.61**
	MAE	0.53	1.03	3.59	1.04	**0.45**
Bed 3	RMSE	0.55	3.93	3.96	1.26	**0.19**
	MAE	0.46	3.71	3.50	1.21	**0.13**
Bed 4	RMSE	0.83	1.36	4.68	1.36	**0.67**
	MAE	0.55	0.99	3.59	1.27	**0.36**
Bed 5	RMSE	0.49	3.94	2.69	1.08	**0.42**
	MAE	0.32	3.71	2.18	0.85	**0.34**

VII. DISCUSSIONS

LSTM+PSO model, especially when utilizing a train-test split ratio of 70:30, is highly effective in predicting results for a difference of beds, namely, Bed 1, Bed 2, Bed 3, Bed 4, and Bed 5. The model can perform well. The prediction result of LSTM+PSO using rolling origin highlights the reliability of the LSTM+PSO approach in achieving precise predictions across different bed types. However, it is worth noting that the rolling origin method, while promising, may utilize only some of the dataset as effectively as a fixed train-test split of 70:30, as it involves random cross-validation that may exclude some data observations [33].

The most notable outcome of this study is that the LSTM+PSO model outperformed other models such as LSTM, SVR, Random Forest, and SARIMAX, as evident by lower RMSE and MAE values. It highlights the effectiveness of the LSTM+PSO model in making predictions on the dataset. The LSTM+PSO can be advantageous in terms of exploration, exploitation [34], [35], stochastic search, optimal capability, and the ability to handle global and local optima [36], [37]. PSO is known for its ability to explore the search space effectively. When combined with LSTM, which tends to converge quickly to local optima, PSO helps explore different regions of the parameter space, potentially leading to better global solutions.

While LSTM is good at fine-tuning, the PSO algorithm attracts other particles toward the region. The stochastic behaviour helps to escape local optima, which can be especially beneficial when combined with LSTM, which tends to be deterministic [38]–[40]. LSTM+PSO can harness the optimal capability of PSO to find the best set of hyperparameters of weights for the LSTM that can lead to improved overall model performance. It also allows for a more robust optimization process to improve model performance, especially in complex and high-dimensional search spaces. Thus, PSO is designed to handle both global and local optima that complement LSTM's ability to fine-tune models to local patterns in small datasets.

VIII. CONCLUSION

The LSTM+PSO method is proposed for NLC prediction of a drier bed, which can assist manual moisture adsorption capacity test. LSTM+PSO utilized a new particle representation to obtain a robust model for predicting outcomes for different bed types. Compared with LSTM, SVR RF, and SARIMAX, the proposed LSTM+PSO performs better, achieving significantly lower RMSE and MAE values for a small dataset. Additionally, a new particle representation LSTM+PSO, an efficient model, is achieved for predicting outcomes on different bed types. This achievement provides significant possibilities for improving future investigations within this domain. In future research efforts, it is suggested to incorporate additional data from alternative sources, such as other beds, together with experimental data collected over a period. This approach will facilitate the analysis of temporal predictions, allowing for the refinement of prediction models and the enhancement of prediction accuracy. Additional techniques, such as cuckoo search and firefly algorithm will be employed in future research to reduce the error and obtain an optimal solution.

ACKNOWLEDGMENT

The authors would like to acknowledge the PETRONAS Academia Collaboration Grant, Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), and Universiti Teknologi MARA for the financial support provided to this research project.

REFERENCES

[1] S. Banooni, E. Hajidavalloo, and M. Dorfeshan, "Experimental and numerical study of the effects of pre-drying of S-PVC using a pneumatic dryer," Powder Technol, vol. 338, pp. 220–232, 2018, doi: 10.1016/j.powtec.2018.06.027.

[2] B. Lan, P. Zhao, J. Xu, B. Zhao, M. Zhai, and J. Wang, "CFD-DEM-IBM simulation of particle drying processes in gas-fluidized beds," Chem Eng Sci, vol. 255, 2022, doi: 10.1016/j.ces.2022.117653.

[3] M. R. Nukulwar and V. B. Tungikar, "Recent development of the solar dryer integrated with thermal energy storage and auxiliary units," Thermal Science and Engineering Progress, vol. 29. Elsevier Ltd, 2022. doi: 10.1016/j.tsep.2021.101192.

[4] W. Su, D. Ma, Z. Lu, W. Jiang, F. Wang, and Z. Xiaosong, "A novel absorption-based enclosed heat pump dryer with combining liquid desiccant dehumidification and mechanical vapor recompression: Case study and performance evaluation," Case Studies in Thermal Engineering, vol. 35, 2022, doi: 10.1016/j.csite.2022.102091.

[5] M. Mohseni, A. Kolomijtschuk, B. Peters, and M. Demoulling, "Biomass drying in a vibrating fluidized bed dryer with a Lagrangian-Eulerian approach," International Journal of Thermal Sciences, vol. 138, pp. 219–234, 2019, doi: 10.1016/j.ijthermalsci.2018.12.038.

[6] M. Djaeni, D. Q. A'yuni, M. Alhanif, C. L. Hii, and A. C. Kumoro, "Air dehumidification with advance adsorptive materials for food drying: A critical assessment for future prospective," Drying Technology, vol. 39, no. 11, pp. 1648–1666, 2021, doi: 10.1080/07373937.2021.1885042.

[7] B. El Fil and S. Garimella, "Energy-efficient gas-fired tumble dryer with adsorption thermal storage," Energy, 2022, doi: 10.1016/j.energy.2021.121708.

[8] A. M. Nandhu Lal et al., "A comparison of the Refrigerated Adsorption Drying of Daucus carota with fluidized bed drying," LWT, 2022, doi: 10.1016/j.lwt.2021.112749.

[9] H. Daghooghi-Mobarakeh, M. Miner, L. Wang, R. Wang, and P. E. Phelan, "Application of ultrasound in regeneration of silica gel for industrial gas drying processes," Drying Technology, vol. 40, no. 11, pp. 2251–2259, 2022, doi: 10.1080/07373937.2021.1929296.

[10] K. Zhang and M. Hong, "Forecasting crude oil price using LSTM neural networks," Data Science in Finance and Economics, 2022, doi: 10.3934/dsfe.2022008.

[11] A. Manowska and A. Bluszcz, "Forecasting Crude Oil Consumption in Poland Based on LSTM Recurrent Neural Network," Energies (Basel), 2022, doi: 10.3390/en15134885.

[12] R. H. Assaad and S. Fayek, "Predicting the Price of Crude Oil and its Fluctuations Using Computational Econometrics: Deep Learning, LSTM, and Convolutional Neural Networks," Econometric Research in Finance, 2021, doi: 10.2478/erfin-2021-0006.

[13] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM," 2019, [Online]. Available: http://arxiv.org/abs/1911.09512

[14] K. Yan, W. Li, Z. Ji, M. Qi, and Y. Du, "A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households," IEEE Access, 2019, doi: 10.1109/ACCESS.2019.2949065.

[15] D. Durand, J. Aguilar, and M. D. R-Moreno, "An Analysis of the Energy Consumption Forecasting Problem in Smart Buildings Using LSTM," Sustainability (Switzerland), 2022, doi: 10.3390/su142013358.

[16] A. A. Ewees, M. A. A. Al-qaness, L. Abualigah, and M. A. Elaziz, "HBO-LSTM: Optimized long short term memory with heap-based optimizer for wind power forecasting," Energy Convers Manag, vol. 268, 2022, doi: 10.1016/j.enconman.2022.116022.

[17] S. U. M. Rao, K. V. Rao, and P. V. G. D. P. Reddy, "Medical Big Data Analysis using LSTM based Co-Learning Model with Whale Optimization Approach," International Journal of Intelligent Engineering and Systems, vol. 15, no. 4, pp. 627 – 636–627 – 636, 2022, doi: 10.22266/ijies2022.0831.56.

[18] Y. Choi et al., "Time-series clustering approach for training data selection of a data-driven predictive model: Application to an industrial bio 2,3-butanediol distillation process," Comput Chem Eng, vol. 161, 2022, doi: 10.1016/j.compchemeng.2022.107758.

[19] R. Gomes Santiago et al., "Investigation of premature aging of zeolites used in the drying of gas streams," Chem Eng Commun, vol. 206, no. 11, pp. 1378–1385, 2019, doi: 10.1080/00986445.2018.1533468.

[20] B. El Fil and S. Garimella, "Heat recovery, adsorption thermal storage, and heat pumping to augment gas-fired tumble dryer efficiency," J Energy Storage, 2022, doi: 10.1016/j.est.2021.103949.

[21] L. A. Nikolaeva, R. Zainullina, and A. K. Al'-Okbi, "Adsorption Drying of Natural Gas by Carbonate Sludge," Chemical and Petroleum Engineering, vol. 54, no. 11–12, pp. 919–925, 2019, doi: 10.1007/s10556-019-00572-2.

[22] M. R. Petryk, A. Khimich, M. M. Petryk, and J. Fraissard, "Experimental and computer simulation studies of dehydration on microporous adsorbent of natural gas used as motor fuel," Fuel, vol. 239, pp. 1324–1330, 2019, doi: 10.1016/j.fuel.2018.10.134.

[23] T. Mołczan and P. Cyklis, "Mathematical Model of Air Dryer Heat Pump Exchangers," Energies (Basel), 2022, doi: 10.3390/en15197092.

[24] S. K. Patel and M. H. Bade, "Energy analysis and heat recovery opportunities in spray dryers applied for effluent management," Energy Convers Manag, 2019, doi: 10.1016/j.enconman.2019.02.065.

[25] A. Rieser, D. Herrera-Avellanosa, E. Leonardi, M. Larcher, and R. Pfluger, "Experimental measurement of material's drying coefficient for internal insulation: New approaches for laboratory testing," in IOP Conference Series: Earth and Environmental Science, 2021. doi: 10.1088/1755-1315/863/1/012048.

[26] V. Meshalkin, V. Bobkov, M. Dli, and V. Dovì, "Optimization of energy and resource efficiency in a multistage drying process of phosphate pellets," Energies (Basel), 2019, doi: 10.3390/en12173376.

[27] L. Xu, Y. Wang, L. Mo, Y. Tang, F. Wang, and C. Li, "The research progress and prospect of data mining methods on corrosion prediction of oil and gas pipelines," Eng Fail Anal, vol. 144, p. 106951, 2023, doi: https://doi.org/10.1016/j.engfailanal.2022.106951.

[28] A. S. B. Karno, "Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory)," Journal of Informatic and Information Security, 2020, doi: 10.31599/jiforty.v1i1.133.

[29] E. Kristiani, H. Lin, J. R. Lin, Y. H. Chuang, C. Y. Huang, and C. T. Yang, "Short-Term Prediction of PM2.5 Using LSTM Deep Learning Methods," Sustainability (Switzerland), 2022, doi: 10.3390/su14042068.

[30] Y. T. Tsan, D. Y. Chen, P. Y. Liu, E. Kristiani, K. L. P. Nguyen, and C. T. Yang, “The Prediction of Influenza-like Illness and Respiratory Disease Using LSTM and ARIMA,” Int J Environ Res Public Health, 2022, doi: 10.3390/ijerph19031858.

[31] S. H. Bhatti, F. W. Khan, M. Irfan, and M. A. Raza, “An effective approach towards efficient estimation of general linear model in case of heteroscedastic errors,” Commun Stat Simul Comput, 2023, doi: 10.1080/03610918.2020.1856874.

[32] İ. Y. Genç, “Prediction of storage time in different seafood based on color values with artificial neural network modeling,” J Food Sci Technol, 2022, doi: 10.1007/s13197-021-05269-0.

[33] J. A. Fiorucci and F. Louzada, “GROEC: Combination method via Generalized Rolling Origin Evaluation,” Int J Forecast, vol. 36, no. 1, pp. 105–109, 2020, doi: https://doi.org/10.1016/j.ijforecast.2019.04.013.

[34] J. H. Elrefaei, A. H. Madian, A. Yahya, M. K. Shaat, and R. M. Fikry, “A Modified Particle Swarm Optimization Approach for Latency of Wireless Sensor Networks,” International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 676–685, 2021, doi: 10.14569/IJACSA.2021.0120679.

[35] M. Elveny, M. K. M. Nasution, and R. B. Y. Syah, “A Hybrid Metaheuristic Model for Efficient Analytical Business Prediction,” International Journal of Advanced Computer Science and Applications, vol. 14, no. 8, pp. 433–440, 2023, doi: 10.14569/IJACSA.2023.0140848.

[36] G. Dominico and R. S. Parpinelli, “Multiple global optima location using differential evolution, clustering, and local search,” Appl Soft Comput, vol. 108, p. 107448, 2021, doi: https://doi.org/10.1016/j.asoc.2021.107448.

[37] Z. Meng, Y. Zhong, G. Mao, and Y. Liang, “PSO-sono: A novel PSO variant for single-objective numerical optimization,” Inf Sci (N Y), vol. 586, pp. 176–191, 2022, doi: https://doi.org/10.1016/j.ins.2021.11.076.

[38] G. Tang, J. Sheng, D. Wang, and S. Men, “Continuous Estimation of Human Upper Limb Joint Angles by Using PSO-LSTM Model,” IEEE Access, 2021, doi: 10.1109/ACCESS.2020.3047828.

[39] X. Gao, Y. Guo, D. A. Hanson, Z. Liu, M. Wang, and T. Zan, “Thermal error prediction of ball screws based on PSO-LSTM,” International Journal of Advanced Manufacturing Technology, 2021, doi: 10.1007/s00170-021-07560-y.

[40] Q. Li, X. Chai, C. Zhang, X. Wang, and W. Ma, “Prediction Model of Ischemic Stroke Recurrence Using PSO-LSTM in Mobile Medical Monitoring System,” Comput Intell Neurosci, 2022, doi: 10.1155/2022/8936103.

A Deep Learning-based Approach for Vision-based Weeds Detection

Yan Wang*
Department of Architecture Engineering, Shijiazhuang College of Applied Technology, Shijiazhuang 050081, China

***Abstract*—Weed detection is an essential component of smart agriculture, and the use of remote sensing technologies has the potential to significantly improve weed management practices, reduce herbicide usage, and increase crop yields. This study proposed an approach to weed detection using computer vision and deep learning technologies. By utilizing remote sensing methods based on DL, this approach has the potential to optimize weed management strategies, minimize herbicide use, and enhance crop productivity. The weed detection algorithm is based on the Yolov8 framework, and a custom model is trained using images from popular datasets as well as the internet. To evaluate the model's effectiveness, it is tested on both validation and testing sets. Furthermore, the model's performance is assessed using images that are not included in the original dataset. As experimental results shown, the deep learning-based approach is a promising solution for weed detection in agriculture.**

Keywords—Smart agriculture; weed detection; remote sensing; deep learning; computer vision

I. Introduction

Smart agriculture, also known as precision agriculture [1], involves the use of technology to enhance and optimize crop production while minimizing environmental impact. This includes using sensors and monitoring systems to collect data on soil conditions, weather patterns, and plant growth, as well as utilizing machine learning algorithms and automation to improve efficiency and reduce waste. Smart agriculture also encompasses the use of drones, robotics, and other advanced technologies to perform tasks such as planting, watering, and harvesting. Overall, smart agriculture aims to increase yields, reduce resource usage, and promote sustainable farming practices [2].

Weeds are unwanted plants that compete with crops for resources such as nutrients, water, and sunlight, which can reduce the yield and quality of agricultural products [3]. Weeds can also serve as hosts for pests and diseases, which can further impact crop health and productivity. To manage weeds, farmers may use various methods such as mechanical cultivation, hand weeding, or chemical herbicides. However, the use of herbicides can have negative effects on the environment and human health, so there is growing interest in alternative weed management strategies such as integrated weed management, cover cropping, and crop rotation. Effective weed management is essential for maintaining healthy and productive agricultural systems while minimizing negative impacts on the environment and human health [4, 5].

Weed detection is a critical aspect of precision agriculture, as it allows farmers to identify and manage weeds more effectively [6]. There are several methods for detecting weeds in agricultural fields, including visual inspections, manual sampling, and remote sensing technologies. Visual inspections involve physically observing the crop and looking for signs of weed growth. This method can be time-consuming and labor-intensive, but it can be useful for identifying small infestations or for crops with lower weed densities. Manual sampling involves collecting samples of soil or plant material from different locations in the field and analyzing them for the presence of weeds [7]. This method is more accurate than visual inspections but can still be time-consuming and requires trained personnel. Remote sensing technologies, such as satellites, drones, or ground-based sensors, can provide rapid and accurate weed detection across large areas [8, 9]. These technologies use various sensors such as multispectral or hyperspectral cameras, which can detect differences in plant color, reflectance, or texture, to identify and map weeds [10]. Machine learning algorithms can then be used to analyze the data and develop weed management strategies [11].

Deep learning is a subset of machine learning that uses artificial neural networks to automatically learn and identify patterns in data [11, 12]. In the context of weed detection, deep learning algorithms can be trained on large datasets of images to recognize and classify different weed species. This involves using convolutional neural networks (CNNs) to extract features from the images and then using a combination of fully connected layers and softmax classifiers to classify the images [13]. Deep learning-based weed detection systems have shown high accuracy rates in detecting and classifying weeds, even in complex agricultural environments [14, 15]. These systems have the potential to revolutionize weed management practices by providing farmers with rapid and accurate information about weed infestations, allowing for more targeted and effective weed control strategies.

In this study, a deep learning-based method is proposed for weed detection. In DL based which is in remote sensing technologies it has the potential to significantly improve weed management practices, reduce herbicide usage, and increase crop yields. In order to detect weeds, a Yolo based algorithm is developed to detect the weeds. For this detection, a model is trained using collected images from internet and other popular dataset. The generated model is evaluated and test using associated validation and testing sets. Finally, the model is tested with images outside of our dataset to make sure the performance of the method is effective.

The main research contributions of this study are as follows:

1) The study introduces a novel deep learning-based method for weed detection in remote sensing technologies, offering the potential to enhance weed management practices and contribute to reductions in herbicide usage, ultimately leading to increased crop yields.

2) A YOLO-based algorithm is developed as part of the research, providing an effective and efficient means of detecting weeds, thereby contributing to the advancement of automated weed detection systems.

3) The research contributes by presenting a meticulously trained model, utilizing a diverse set of images collected from the internet and other popular datasets, and evaluates its performance not only on associated validation and testing sets but also on external images, ensuring the method's effectiveness beyond the initial dataset.

II. Related Works

In research [12], custom lightweight deep learning models are suggested for detecting weeds in soybean crops. The models were trained using a dataset of images depicting soybean crops with varying weed types. The findings demonstrate that the proposed models outperform conventional machine learning algorithms in terms of speed, memory usage, and accuracy. The authors suggest that the custom lightweight deep learning models can efficiently detect weeds in soybean crops, which can result in improved crop management and reduced usage of herbicides.

Peng et al, [16] presented an enhanced RetinaNet network for detecting weeds in paddy fields. The proposed network employs residual connections and feature pyramid network (FPN) to improve the accuracy of weed detection. The dataset used in this study consists of images of paddy fields containing different types of weeds, which were used for both training and testing the network. The findings of the study indicate that the proposed network outperforms existing methods and is effective in detecting weeds in paddy fields. The authors suggest that the improved RetinaNet network has the potential to aid in weed management and reduce the need for herbicides in agriculture.

Haq et al, [17] developed an automated weed detection system that relies on CNNs and UAV imagery. The proposed system captures aerial images of crop fields using UAVs and employs CNNs to differentiate between crops and weeds. The CNNs are trained using a dataset of images containing both crops and weeds. The study reveals that the system can accurately detect weeds in a timely manner. Their experimental results show using UAVs and CNNs for weed detection can lead to better weed management in agriculture, enhancing efficiency and accuracy.

The authors in study [18] presented an enhanced version of the YOLO v4 algorithm for detecting weeds in images of carrot fields. The proposed algorithm leverages data augmentation and transfer learning techniques to improve the performance of the YOLO v4 model. The authors collected a dataset of images of carrot fields with and without weeds, and the algorithm was trained and evaluated on this dataset. The findings indicate that the improved YOLO v4 algorithm surpasses the traditional YOLO v4 and other advanced algorithms. The authors suggest that the algorithm can effectively identify weeds in carrot fields, contributing to weed management and minimizing the use of herbicides. The study highlights the potential of deep learning methods in agriculture for weed detection.

Alam [5] proposed a machine-learning based system for real-time crop/weed detection and classification, facilitating variable-rate spraying in precision agriculture. The system employs a CNN to detect and differentiate crops and weeds based on their visual features. The study reveals that the proposed system accurately identifies crops and weeds, which can improve targeted spraying and minimize herbicide usage. This study showed that the machine-learning based approach can enhance precision agriculture and promote sustainable crop management practices.

III. Methodology

A. *Yolov8 Algorithm*

The YOLOv8 is an efficient object detection model that was introduced in early of 2023 [19]. It is an improvement over the previous versions of YOLO, which are known for their speed and accuracy in object detection. YOLOv8 is designed to be more accurate than its predecessors while still maintaining real-time performance. The YOLOv8 architecture is composed of several components, including a backbone network, a neck network, and a head network. Fig. 1 shows the architecture of YOLOv8 network.

The backbone network is responsible for extracting features from the input image, while the neck network and the head network are responsible for detecting objects and generating bounding boxes. It uses several equations and formulas in its implementation, including:

1) Sigmoid function: YOLOv8 uses a sigmoid function to transform the predicted outputs into probabilities. The sigmoid function is defined as follows:

$$\text{sigmoid}(x) = 1 / (1 + e\text{^}{-}x) \quad (1)$$

where, *x* is the input to the function.

2) Intersection over Union (IoU): IoU is used to measure the overlap between two bounding boxes. It is defined as the ratio of the area of the intersection of the two bounding boxes to the area of their union.

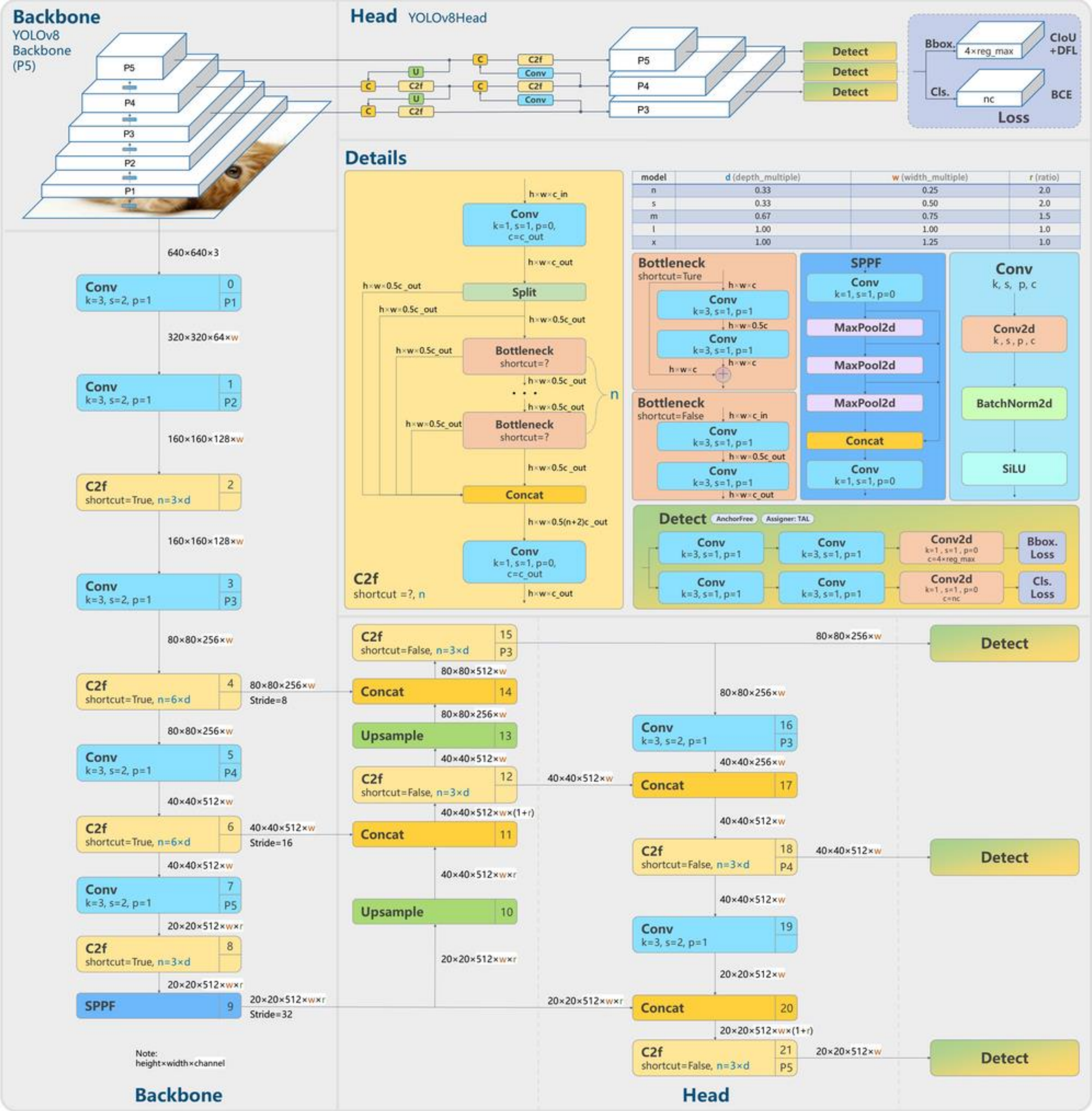
$$\text{IoU}(A, B) = (\text{area of intersection between A and B}) / (\text{area of union between A and B}) \quad (2)$$

3) Anchor boxes: YOLOv8 uses anchor boxes to predict the location and size of objects in the image. Anchor boxes are fixed bounding boxes of different sizes and aspect ratios that are placed at various locations in the image.

4) Loss function: The loss function used in YOLOv8 is a combination of three different losses: the localization loss, the confidence loss, and the class loss. The localization loss measures the difference between the predicted and ground-truth bounding box coordinates. The confidence loss measures the difference between the predicted and ground-truth objectness scores. The class loss measures the difference between the predicted and ground-truth class probabilities.

5) YOLOv8 output tensor: The output of YOLOv8 is a tensor that contains predictions for each anchor box. Each anchor box has a corresponding set of predicted values, which include the class probabilities, objectness score, and bounding box coordinates. The tensor is typically represented as follows:

[batch_size, grid_size, grid_size, num_anchors, num_classes + 5] (3)



model	d (depth_multiple)	w (width_multiple)	r (ratio)
n	0.33	0.25	2.0
s	0.33	0.50	2.0
m	0.67	0.75	1.5
l	1.00	1.00	1.0
x	1.00	1.25	1.0

Fig. 1. Architecture of YOLOv8 network [20].

B. Dataset

In this study, a dataset is used from internet resource. The dataset includes images taken from a public dataset in Roboflow. We have totally 4239 images in the dataset. Among these images, augmentation process is performed for extending the dataset. The structure for training task from this dataset, 87% or 3700 images for training set, 9% or 359 images for validation set, and 4% or 180 images for testing set are organized. Some images from the dataset are shown in Fig. 2.

C. Google Colab

To conduct our experiments, we utilized Google's Colab research platform, which offers access to high-performance GPUs at no cost. We conducted all of our training and testing on a 12GB NVIDIA Tesla T4 GPU, which is described in more detail in Fig. 3. Our models were trained with a maximum of 2500 iterations, a batch size of four images, and an image size of 640.

D. Comparison of Yolo Models

This section presents a companion of different models of Yolo networks, the purpose of this comparison is to justify why Yolov8 is selected in this study. Based on published performance analysis of different Yolo based models [20], this investigation is conducted. For this investigation, we can analyze the model’s graph where the X-axis represents the mean average precision (mAP) percentage, and the Y-axis represents the number of parameters in each YOLO-based model. The graph displays curves corresponding to different YOLO model versions: YOLOv8, YOLOv7, YOLOv6, and YOLOv5. Fig. 3 shows this graph.

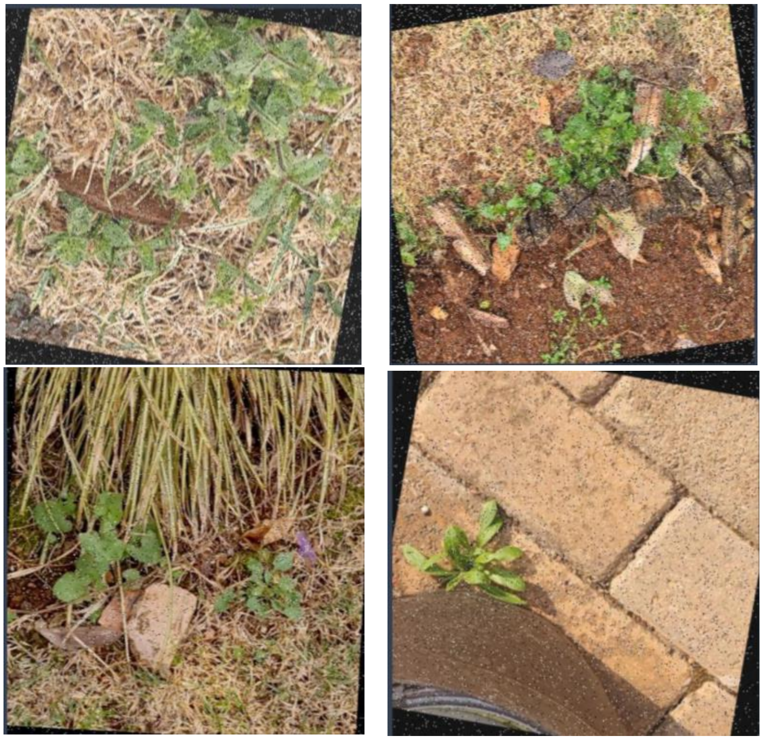


Fig. 2. Sample images of the dataset [20].

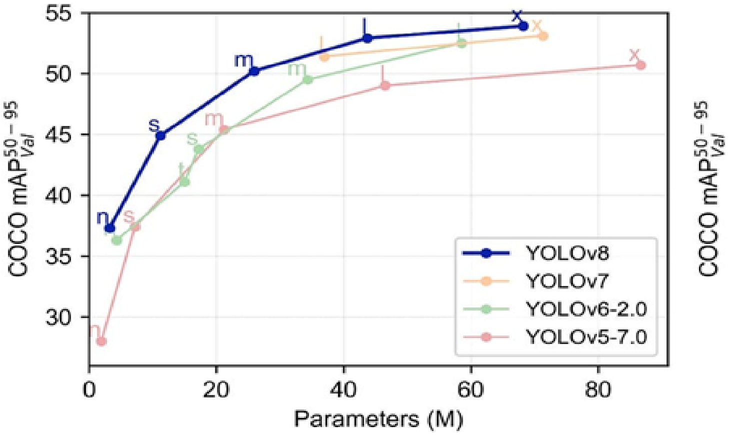


Fig. 3. Comparison of Yolo models in terms of mAP [20].

The graph depicts a comparison between the mAP percentages and the number of parameters for each YOLO-based model. The curves demonstrate the trade-off between the mAP performance and the complexity of the models, represented by the number of parameters. YOLOv8, having the best mAP performance among the models, exhibits a curve that consistently outperforms the other models in terms of mAP percentage. This indicates that YOLOv8 achieves higher accuracy without excessively increasing model complexity, making it a more efficient and scalable choice. Therefore, based on the assumptions provided, the graph illustrates that YOLOv8 surpasses the other models in terms of mAP performance while maintaining a reasonable number of parameters. This makes YOLOv8 the preferred option among the YOLO-based models considered in the graph, as it offers superior accuracy without excessive model complexity.

Moreover, we can analyze the graph where the X-axis represents the average precision (AP) percentage, and the Y-axis represents the performance of PyTorch FP16 running on the RTX 3080 platform. The performance measurements are conducted on the COCO dataset. The graph includes curves corresponding to different YOLO-based model versions: YOLOv8, YOLOv8-seg, YOLOv7, YOLOv6, YOLOv6, and YOLOv5.

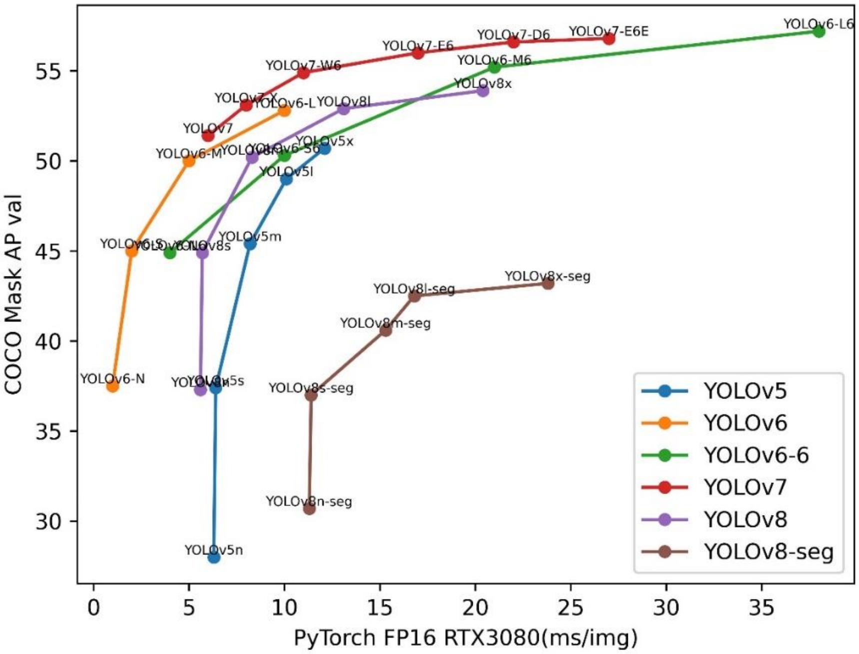


Fig. 4. Comparison of performance for Yolo models in terms of AP [20].

As shown in Fig. 4, the graph presents a comparison between the AP percentages and the performance of PyTorch FP16 running on the RTX 3080 platform for each YOLO-based model. The curves showcase the relationship between AP performance and the computational efficiency of the models. YOLOv8-seg, being the model with the best AP performance according to the assumption, exhibits a curve that consistently outperforms the other models in terms of AP percentage. This indicates that YOLOv8-seg achieves higher accuracy in object detection on the COCO dataset compared to the other versions.

The YOLOv8-seg's superiority is justified not only in terms of AP but also in relation to the computational efficiency represented by the PyTorch FP16 performance on the RTX 3080 platform. Despite its superior AP performance, YOLOv8-seg manages to maintain efficient performance on the given hardware platform, suggesting that it strikes a balance between accuracy and computational efficiency. Therefore, based on the assumptions provided, the graph illustrates that YOLOv8-seg outperforms the other models in terms of AP performance on the COCO dataset while maintaining efficient performance on the specified hardware platform. This makes YOLOv8-seg the preferred choice among the YOLO-based models considered in the graph, as it offers higher accuracy without compromising computational efficiency.

IV. Experimental Results

This section presents experimental results and discuss about performance evaluation in details. Firstly, experimental result is presented from the trained model using above details, and then performance evaluation is discussed. Fig. 5 shows some experimental results.

Performance evaluation is an essential step in the development of object detection models, including the YOLO object detection algorithm. Model evaluation helps assess the performance of a model and determine whether it is meeting the desired accuracy criteria. Popular performance metrics used for model evaluation in object detection tasks are precision, recall, and mAP (mean Average Precision). Fig. 6 shows the performance results.

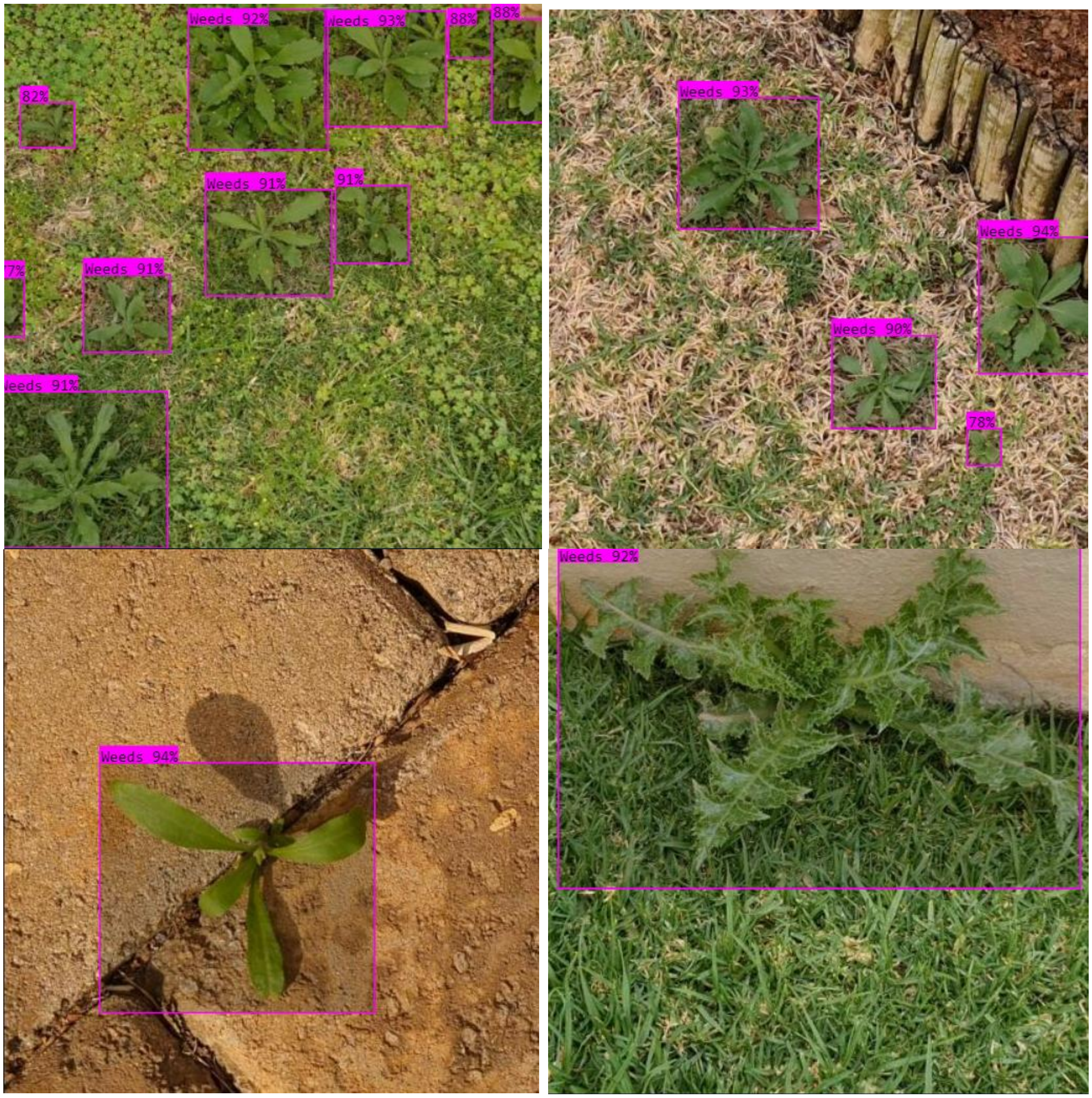


Fig. 5. The result of our experiments.

To present the statistical reporting and data presentation in evaluating the YOLOv8-based model for weed detection, this study provides a detailed breakdown of precision, recall, and mAP values. Instead of a single mAP value, presenting precision-recall curves at different confidence thresholds can offer a nuanced understanding of the model's performance across a range of decision-making points. This not only adds depth to the analysis but also provides insights into the trade-off between precision and recall, aiding in decision-making for real-world applications. Additionally, including a confusion matrix or a similar visual representation would offer a more granular view of the model's strengths and weaknesses, particularly in terms of false positives and false negatives.

As shown in Fig. 6, precision, recall, and mAP are essential metrics used to assess the performance of a model, including its effectiveness in weed detection:

Precision: Precision measures how well the model predicts true positive instances while minimizing false positives. High precision indicates that the model is accurate in identifying weeds and has fewer false alarms.

Recall: Recall measures the model's ability to correctly identify all positive instances, or in this case, accurately detecting weeds. High recall suggests that the model is effective at capturing most of the weeds present in the dataset.

mAP: mAP provides a comprehensive evaluation of the model's performance by considering both precision and recall at various thresholds. A higher mAP indicates a more accurate and effective model for weed detection.

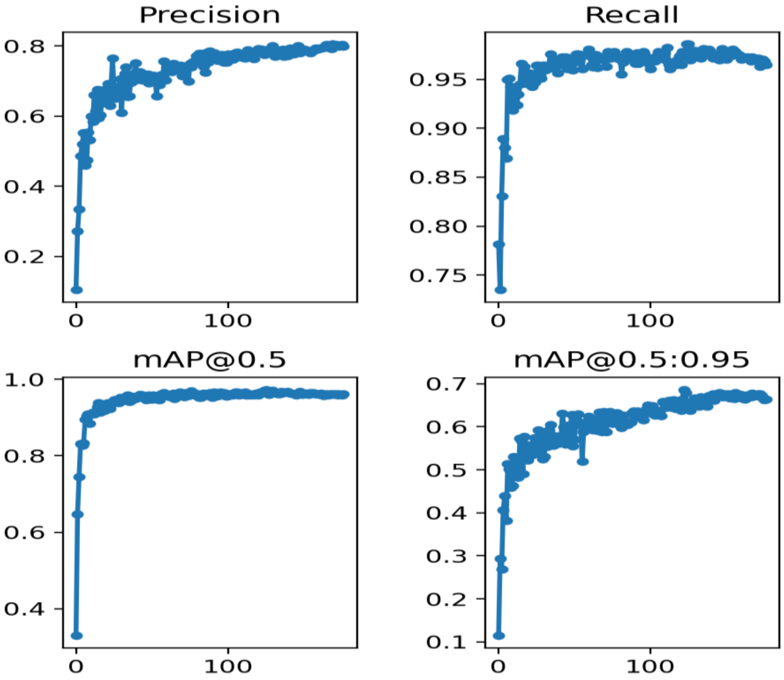


Fig. 6. The performance results.

By considering the precision, recall, and mAP metrics, we can assess the effectiveness of the generated YOLOv8 model for weed detection. If the precision curve demonstrates high precision values across different thresholds, it indicates that the model reliably detects weeds while minimizing false positives. A steep rise in precision suggests the model is precise at differentiating between weeds and non-weed instances. Similarly, if the recall curve shows high recall values, it implies that the model successfully captures a significant number of weed instances, reducing the chances of missing any weeds. Lastly, a high mAP indicates a balance between precision and recall, indicating that the model achieves both accurate weed detection and minimizes false positives.

Therefore, by evaluating the precision, recall, and mAP metrics and ensuring high values for all these measures, we found that the generated YOLOv8 model is effective for accurate weed detection.

V. Conclusion

In this research, a deep learning approach for vision-based weeds detection in agriculture is proposed. The algorithm used for weed detection is built on the Yolov8 framework, and a customized model is created by training it on images from popular datasets as well as the internet. To assess the effectiveness of the model, it is tested on both validation and testing datasets, and its performance is evaluated using images that are not part of the original dataset. The experimental findings demonstrate that the deep learning-based approach is a promising solution for detecting weeds in agriculture. However, in this research for limitation addressing purpose, the diversity of the training data sources used for creating the YOLOv8-based weed detection model. While the model is trained on images from popular datasets and the internet, the potential presence of biases in these sources may affect the model's generalizability to a broader range of real-world agricultural scenarios. The use of internet-collected images might introduce variations in terms of lighting conditions, field types, and weed species that are not fully represented in the training dataset. Consequently, the model's performance might be over-optimized for the specific characteristics of the training data, limiting its effectiveness in more diverse and unpredictable agricultural environments. To address this limitation, future studies should focus on systematically expanding the diversity of the training dataset to ensure the model's robustness across a broader spectrum of agricultural conditions. This could involve incorporating images from geographically diverse locations, different seasons, and various agricultural practices. Additionally, efforts should be made to include images that capture the inherent variability in weed species and growth stages, ensuring that the model can accurately detect weeds under a wide range of circumstances. By addressing this limitation, researchers can enhance the model's applicability and reliability in real-world agricultural settings, ultimately contributing to the successful deployment of the proposed deep learning approach on a larger scale.

References

[1] I. Cisternas, I. Velásquez, A. Caro, and A. Rodríguez, "Systematic literature review of implementations of precision agriculture," Computers and Electronics in Agriculture, vol. 176, p. 105626, 2020.

[2] D. C. Tsouros, S. Bibi, and P. G. Sarigiannidis, "A review on UAV-based applications for precision agriculture," Information, vol. 10, no. 11, p. 349, 2019.

[3] S. Shanmugam, E. Assunção, R. Mesquita, A. Veiros, and P. D. Gaspar, "Automated weed detection systems: A review," KnE Engineering, pp. 271–284-271–284, 2020.

[4] S. Kulkarni, S. Angadi, and V. Belagavi, "IoT based weed detection using image processing and CNN," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 3, pp. 606-609, 2019.

[5] M. Alam, M. S. Alam, M. Roman, M. Tufail, M. U. Khan, and M. T. Khan, "Real-time machine-learning based crop/weed detection and classification for variable-rate spraying in precision agriculture," in 2020 7th International Conference on Electrical and Electronics Engineering (ICEEE), 2020: IEEE, pp. 273-280.

[6] B. Liu and R. Bruch, "Weed detection for selective spraying: a review," Current Robotics Reports, vol. 1, pp. 19-26, 2020.

[7] N. Islam et al., "Early weed detection using image processing and machine learning techniques in an Australian chilli farm," Agriculture, vol. 11, no. 5, p. 387, 2021.

[8] A. Wang, W. Zhang, and X. Wei, "A review on weed detection using ground-based machine vision and image processing techniques," Computers and electronics in agriculture, vol. 158, pp. 226-240, 2019.

[9] M. D. Bah, E. Dericquebourg, A. Hafiane, and R. Canals, "Deep learning based classification system for identifying weeds using high-resolution UAV imagery," in Intelligent Computing: Proceedings of the 2018 Computing Conference, Volume 2, 2019: Springer, pp. 176-187.

[10] Z. Wu, Y. Chen, B. Zhao, X. Kang, and Y. Ding, "Review of weed detection methods based on computer vision," Sensors, vol. 21, no. 11, p. 3647, 2021.

[11] A. M. Hasan, F. Sohel, D. Diepeveen, H. Laga, and M. G. Jones, "A survey of deep learning techniques for weed detection from images," Computers and Electronics in Agriculture, vol. 184, p. 106067, 2021.

[12] N. Razfar, J. True, R. Bassiouny, V. Venkatesh, and R. Kashef, "Weed detection in soybean crops using custom lightweight deep learning models," Journal of Agriculture and Food Research, vol. 8, p. 100308, 2022.

[13] D. Patel, M. Gandhi, H. Shankaranarayanan, and A. D. Darji, "Design of an Autonomous Agriculture Robot for Real-Time Weed Detection Using CNN," in Advances in VLSI and Embedded Systems: Select Proceedings of AVES 2021: Springer, 2022, pp. 141-161.

[14] C. T. Selvi, R. S. Subramanian, and R. Ramachandran, "Weed Detection in Agricultural fields using Deep Learning Process," in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, vol. 1: IEEE, pp. 1470-1473.

[15] A. Nasiri, M. Omid, A. Taheri-Garavand, and A. Jafari, "Deep learning-based precision agriculture through weed recognition in sugar beet fields," Sustainable Computing: Informatics and Systems, vol. 35, p. 100759, 2022.

[16] H. Peng, Z. Li, Z. Zhou, and Y. Shao, "Weed detection in paddy field using an improved RetinaNet network," Computers and Electronics in Agriculture, vol. 199, p. 107179, 2022.

[17] M. A. Haq, "CNN Based Automated Weed Detection System Using UAV Imagery," Comput. Syst. Sci. Eng., vol. 42, no. 2, pp. 837-849, 2022.

[18] B. Ying, Y. Xu, S. Zhang, Y. Shi, and L. Liu, "Weed Detection in Images of Carrot Fields Based on Improved YOLO v4," Traitement du Signal, vol. 38, no. 2, 2021.

[19] C. H. Kang and S. Y. Kim, "Real-time object detection and segmentation technology: an analysis of the YOLO algorithm," JMST Advances, pp. 1-8, 2023.

[20] K. Chen et al., "MMDetection: Open mmlab detection toolbox and benchmark," arXiv preprint arXiv:1906.07155, 2019. [Online]. Available: https://github.com/open-mmlab/mmyolo/tree/dev/configs/yolov8.

Detection of Fruit using YOLOv8-based Single Stage Detectors

Xiuyan GAO, Yanmin ZHANG*
Hebei Software Institute, Baoding 071000, China

***Abstract*—In the agricultural sector, the precise detection of fruits plays a pivotal role in optimizing harvesting procedures, minimizing waste, and ensuring the delivery of high-quality produce. Deep learning methods have consistently exhibited superior accuracy compared to alternative techniques, making them a focal point in fruit detection research. However, the ongoing challenge lies in meeting the stringent accuracy requirements essential for real-world applications in agriculture. Addressing this critical concern, this study proposes an innovative solution utilizing the Yolov8 architecture for fruit detection. The methodology involves the meticulous creation of a custom dataset tailored to capture the diverse characteristics of agricultural fruits, followed by rigorous training, validation, and testing processes. Through extensive experimentation and performance evaluations, the findings underscore the exceptional accuracy achieved by the Yolov8-based model. This methodology not only surpasses existing benchmarks but also establishes a robust foundation for transforming fruit detection practices in agriculture. By effectively addressing the challenges associated with accuracy rates, this approach opens new avenues for optimized harvesting, waste reduction, and enhanced efficiency in agricultural practices, contributing significantly to the evolution of precision farming technologies.**

Keywords—Fruit detection; agricultural sector; deep learning; YOLOv8 model; precision agriculture

I. INTRODUCTION

Fruit detection in agriculture is a vital aspect of modern farming practices [1], [2]. The ability to accurately and efficiently identify and assess the ripeness of fruits plays a pivotal role in optimizing agricultural operations, enhancing crop yield, and ensuring the quality of produce [2]. In recent years, the development and utilization of fruit detection technologies have garnered significant attention due to their potential to revolutionize the agricultural industry [3]. The sample of the fruits is depicted in Fig. 1.

The importance of fruit detection in agriculture cannot be overstated. Timely and precise fruit detection aids farmers in optimizing harvesting schedules, reducing waste, and maximizing crop yields. Additionally, it facilitates early detection of diseases and pests, enabling timely intervention and preventing the spread of infestations, ultimately improving the overall health of fruit-bearing plants [4], [5].

While traditional methods of fruit detection have been employed for decades, recent advancements in technology have opened new avenues for enhancing accuracy and efficiency. This research paper explores the latest developments and innovations in fruit detection methods, particularly focusing on the application of deep learning techniques [2], [6].

Existing technologies have made significant strides in fruit detection [4], [7], but deep learning-based approaches have gained prominence among researchers [8], [9]. This shift is primarily attributed to the remarkable capabilities of deep neural networks in handling complex image data. In the subsequent sections [10], [11], it will delve into the current limitations and challenges associated with deep learning-based fruit detection methods, highlighting the need for further research and innovation.

One of the primary motivations for this study is to address the existing limitations and research challenges associated with deep learning-based fruit detection, especially in meeting the high accuracy requirements demanded by the agricultural industry. Achieving precision and reliability in fruit detection is crucial for optimizing harvesting processes and ensuring the quality of the produce. Therefore, it is imperative to explore novel solutions to tackle these challenges and meet the rigorous standards set by the agricultural sector.

This study proposes a deep learning method utilizing Convolutional Neural Networks (CNNs) to address the demanding requirements of fruit detection in agriculture. It adopts a custom dataset and employs rigorous training, validation, and testing processes to develop a robust and efficient model. This approach is founded on the belief that deep learning can provide the accuracy and reliability needed for effective fruit detection in agricultural settings.

This research paper presents three key contributions. First, it generates a custom dataset tailored specifically for fruit detection challenges, providing a valuable resource for future research in this domain. Second, it proposes an efficient deep-learning method that not only detects fruits but also addresses disease detection within the same framework, further enhancing the utility of the model. Third, extensive experiments and performance evaluations are conducted to validate the effectiveness of our method, providing empirical evidence of its potential to revolutionize fruit detection practices in agriculture.

Fig. 1. Sample of fruit images.

II. Relevant Studies

Machine learning and deep learning methods have made substantial contributions to the progress of agricultural sectors, specifically in the fields of disease prediction, classification, and the recognition of fruit types and diseases. These approaches provide a reliable, cost-effective, and swift means of identifying and diagnosing fruit ailments in a non-invasive manner. Numerous scientists have devoted their expertise to the study of fruit detection. Eminent researchers who have made significant strides in this domain encompass:

The paper in [12] presented the Lightweight SM-YOLOv5 algorithm for tomato fruit detection in plant factories. The method employs a modified YOLOv5 architecture optimized for resource-efficient tomato detection. It achieves high accuracy while remaining computationally lightweight. However, a notable limitation is its specialization for tomato detection, which may limit its applicability to other fruits or plant types. Additionally, the paper lacks extensive experimentation and validation on various datasets and real-world conditions. Nonetheless, the Lightweight SM-YOLOv5 offers a promising approach for efficient tomato fruit detection within plant factory environments.

The authors in [13] focused on pineapple fruit detection and localization in natural environments using binocular stereo vision and an improved YOLOv3 model. The method combines depth information from stereo vision with the YOLOv3 model for accurate pineapple detection. However, it is limited in its applicability primarily to pineapple detection and may not generalize well to other fruits or scenarios. Additionally, the paper lacks extensive validation of a wide variety of natural environments and conditions. Nevertheless, the approach represents a promising advancement for pineapple fruit detection in natural settings, showcasing the potential of combining computer vision and deep learning techniques.

The paper in [14] introduces a cherry fruit detection algorithm using an enhanced YOLO-v4 model. The method leverages the YOLO-v4 architecture to achieve accurate cherry detection in images. However, its primary limitation is its specificity to cherry fruit detection, potentially lacking versatility for other fruit types. The paper could benefit from a broader evaluation across different cherry varieties and environmental conditions. Nonetheless, the use of the improved YOLO-v4 model shows promise for enhancing cherry fruit detection precision, offering valuable insights into fruit detection techniques within the agricultural domain.

These authors in [15] introduced a dragon fruit-picking detection method that combines YOLOv7 and PSP-Ellipse. YOLOv7 is employed for object detection, while PSP-Ellipse enhances the accuracy of dragon fruit recognition. However, the limitation lies in its specialization for dragon fruit detection, potentially limiting its applicability to other fruits or objects. Further validation in diverse environmental conditions and fruit varieties would strengthen the method's robustness and utility in agricultural settings. Nonetheless, this approach demonstrates the potential to improve dragon fruit picking efficiency through advanced object detection techniques.

This paper in [16] focused on fruit maturity stage detection and yield estimation in wild blueberries through the use of deep learning Convolutional Neural Networks (CNNs). The method employs CNNs to analyze images of wild blueberry plants, determining fruit maturity and estimating yield. A limitation of the study is that it may require substantial labeled data for training, which can be resource-intensive. Additionally, the model's generalization to different environments and wild blueberry varieties may require further investigation. Nevertheless, the approach demonstrates the potential for improving wild blueberry farming practices through deep learning-based fruit assessment and yield estimation.

The authors in [17] presented a method for detecting tomato plant phenotyping traits using YOLOv5-based single-stage detectors. This approach utilizes YOLOv5 to identify and characterize various traits of tomato plants, facilitating phenotypic analysis. However, the limitation lies in the model's potential sensitivity to variations in environmental conditions, which may affect detection accuracy. Additionally, broader validation across diverse tomato varieties and growth stages could enhance the method's generalization. Nevertheless, this technique showcases promise in automating plant phenotyping tasks, offering valuable insights for agricultural research and crop improvement. The mentioned papers primarily focus on different aspects of fruit detection and utilize various deep-learning models for this purpose. The [12] concentrates on tomato fruit detection in controlled environments, emphasizing the need for efficiency. On the other hand, [13] extends the scope to outdoor settings and employs a stereo vision-based YOLOv3 model.

Similarly, research in [14] narrows its focus to cherry fruit detection and utilizes the YOLOv4 model. In contrast, research in [15] explores dragon fruit detection using YOLOv7 and elliptical detection techniques. The study in [16] extends the application to the maturity stage and yield estimation in wild blueberries, leveraging convolutional neural networks. Lastly, [17] adopts YOLOv5 to identify tomato plant phenotypic traits.

The current literature on fruit detection using deep learning models has made notable strides in achieving high accuracy for specific fruit types in controlled and natural environments. However, a significant gap exists in the lack of algorithms that are both accurate and computationally efficient across a diverse range of fruits and environmental conditions. While individual studies, such as the Lightweight SM-YOLOv5 for tomatoes, binocular stereo vision for pineapples, YOLOv4 for cherries, and YOLOv7 for dragon fruit, demonstrate advancements within their respective domains, their specificity to a single fruit type hampers their widespread applicability. Additionally, the resource-intensive nature of labeled data for training and limited generalization to various fruit varieties and environmental conditions pose challenges. There is a critical need for future research to prioritize the development of algorithms that not only enhance accuracy but also minimize computation costs, enabling real-time processing and practical applications in diverse agricultural settings. Addressing this gap will significantly advance fruit detection techniques in agriculture.

III. Materials and Method

A. Data Collection

This study leveraged a dataset sourced from Roboflow resources [18] to facilitate the fruit detection research. The dataset comprises a diverse collection of fruit images, which served as the foundation for training and evaluating the model.

The dataset encompasses a diverse collection of fruit images, introducing a rich spectrum of variations in terms of size, shape, color, and environmental conditions. This deliberate inclusion of diverse instances is crucial for proving the scalability of our proposed model. Scalability, in the context of the study, refers to the model's ability to generalize effectively across a broad range of scenarios and conditions. By incorporating a wide variety of fruits and their respective attributes, the dataset from Roboflow ensures that the model is exposed to a comprehensive representation of real-world conditions.

Class balance in a dataset refers to the distribution of samples across different classes or categories. In the context of the dataset, class balance assesses whether the number of images for each class is relatively even or if there is a significant imbalance. A balanced dataset ideally has roughly the same number of samples for each class. Looking at the list of classes in the dataset, it appears that there is a varying degree of class balance. Some classes like "apple," "banana," "carrot," "cucumber," "okra," "potato," "sweet-potato," "tomato," and "un-usable" represent specific food items and may have a more balanced distribution if there are a similar number of images for each. However, classes like "Fresh Oranges," "Papaya Fresh," "Rotten," "Rotten Oranges," "bad," "fresh-20%," "fresh-70%," "fresh-90%," and "good" seem to describe the condition or quality of the food items. It's important to consider class balance when training machine learning models because an imbalance can lead to biased predictions, where the model may perform well on the majority class but poorly on minority classes. Fig. 2 shows the class balance of the dataset.

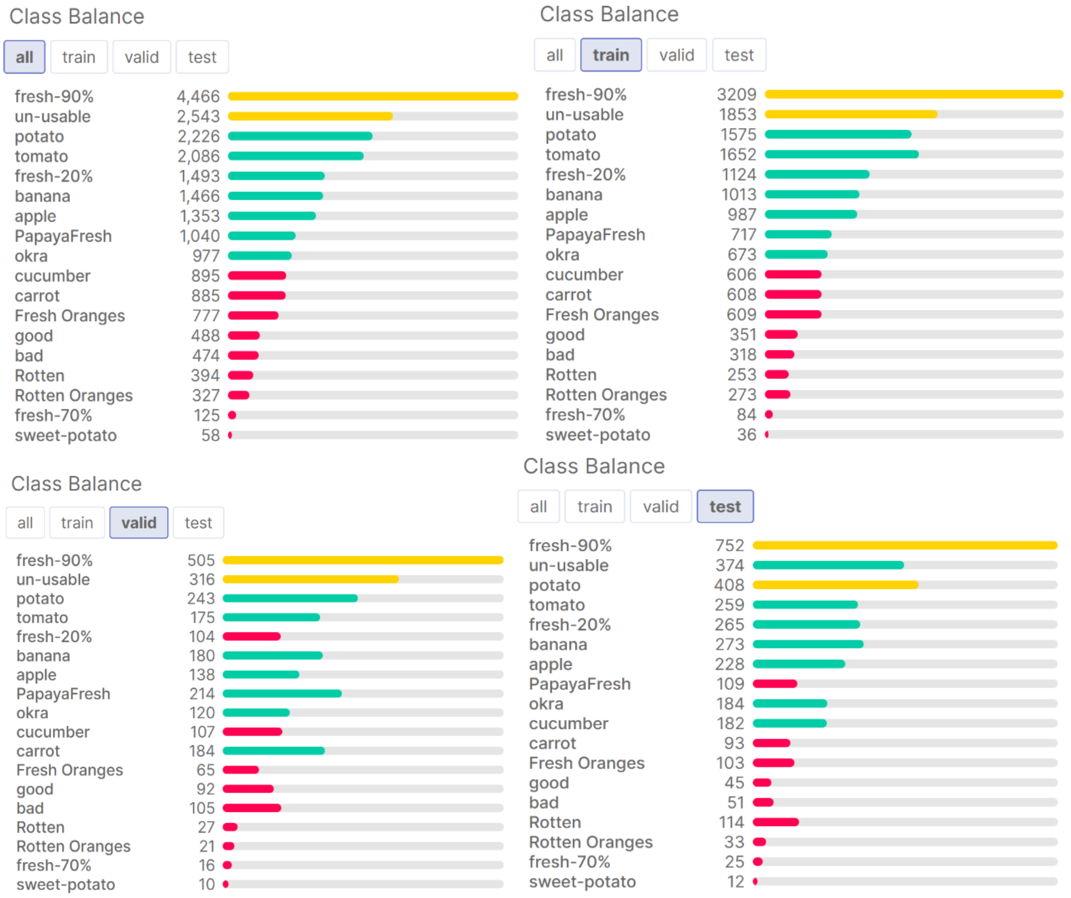
B. Model Training using YOLOv8-based Single Stage Detector

YOLOv8 is indeed a single-stage object detection model. Single-stage detectors (see Fig. 3), in the information of object detection in computer vision, are designed to perform object localization and classification in a single pass through the neural network without the need for a separate region proposal step. The YOLOv8 achieves this as a single-stage detector:

1) Grid-based detection: YOLOv8 divides the input image into a grid, where each grid cell is responsible for predicting objects within its boundaries. The model then predicts bounding boxes (rectangular regions) for objects within each grid cell. This grid-based approach simplifies the object detection process.

2) Multi-scale detection: YOLOv8 uses multiple detection scales to capture objects of different sizes in the same pass. This allows the model to efficiently handle a variety of object scales within the input image.

3) High-speed inference: Being a single-stage detector, YOLOv8 is known for its real-time or near real-time inference capabilities, making it suitable for applications that require fast and accurate object detection, such as autonomous vehicles, surveillance, and robotics. In summary, YOLOv8 is a single-stage object detector that excels in rapid and accurate object detection tasks by directly predicting object bounding boxes and classifications in a single forward pass through the neural network. This efficiency is a key reason for its popularity in various computer vision applications.



Class Balance

all train valid test

fresh-90%	4,466
un-usable	2,543
potato	2,226
tomato	2,086
fresh-20%	1,493
banana	1,466
apple	1,353
PapayaFresh	1,040
okra	977
cucumber	895
carrot	885
Fresh Oranges	777
good	488
bad	474
Rotten	394
Rotten Oranges	327
fresh-70%	125
sweet-potato	58

Class Balance

all **train** valid test

fresh-90%	3209
un-usable	1853
potato	1575
tomato	1652
fresh-20%	1124
banana	1013
apple	987
PapayaFresh	717
okra	673
cucumber	606
carrot	608
Fresh Oranges	609
good	351
bad	318
Rotten	253
Rotten Oranges	273
fresh-70%	84
sweet-potato	36

Class Balance

all train **valid** test

fresh-90%	505
un-usable	316
potato	243
tomato	175
fresh-20%	104
banana	180
apple	138
PapayaFresh	214
okra	120
cucumber	107
carrot	184
Fresh Oranges	65
good	92
bad	105
Rotten	27
Rotten Oranges	21
fresh-70%	16
sweet-potato	10

Class Balance

all train valid **test**

fresh-90%	752
un-usable	374
potato	408
tomato	259
fresh-20%	265
banana	273
apple	228
PapayaFresh	109
okra	184
cucumber	182
carrot	93
Fresh Oranges	103
good	45
bad	51
Rotten	114
Rotten Oranges	33
fresh-70%	25
sweet-potato	12

Fig. 2. The class balance of the dataset.

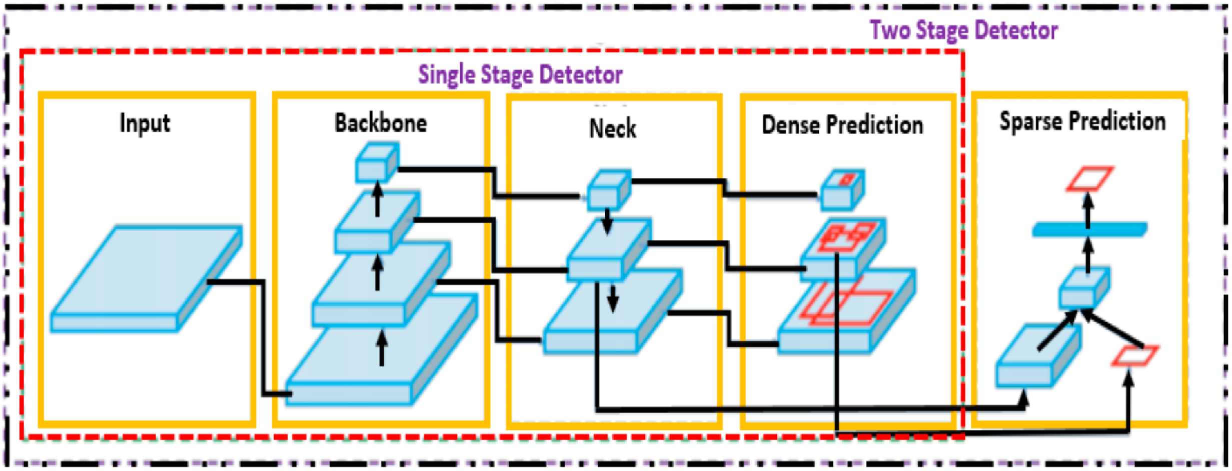


Fig. 3. Object detector anatomy.

C. Model Evaluation Techniques

In the framework of evaluating the performance of a YOLOv8 model for fruit detection, several key metrics are commonly used: F1 score, precision, recall, and mAP (mean Average Precision). These metrics provide valuable insights into the model's ability to detect and classify fruit objects in images accurately. Precision measures the accuracy of the model's positive predictions and the ability to identify fruits correctly. To compute precision, the number of true positive predictions (correctly identified fruits) divide by the total number of positive predictions (true positives plus false positives). High precision indicates that when the model predicts a fruit, it is usually accurate.

$$\text{Precision} = TP / (TP + FP)$$

Recall assesses the model's capability to find all the actual positive instances, i.e., fruits in this case. It calculates the ratio of true positive predictions to the total number of actual positive instances. High recall implies that the model can successfully detect most of the fruits present.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

The F1 score is the harmonic mean of precision and recall. It provides a balanced evaluation of both false positives and false negatives. The F1 score is particularly useful when it considers both precision and recall simultaneously. A high F1 score suggests a model with good overall performance.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

mAP @0.5 is a comprehensive metric widely used in object detection tasks, including fruit detection. It quantifies the precision-recall trade-off across different confidence thresholds for object detection. mAP calculates the area under the precision-recall curve, providing a holistic assessment of the model's performance at varying confidence levels. Higher mAP indicates better overall detection accuracy. Calculate precision, recall, and F1 score to assess the model's accuracy and ability to balance true positives, false positives, and false negatives. These metrics collectively provide valuable insights into how well the YOLOv8 model is performing in fruit detection and help in making informed decisions for model refinement and optimization.

IV. Results

A. Model Evaluation for YOLOv8s

Precision, recall, precision-confidence, and F1 score curves are vital for evaluating the efficiency of a YOLOv8s model in fruit detection. Precision measures the accuracy of positive predictions, recall gauges the model's ability to capture actual instances, precision-confidence reflects the trade-off between confidence thresholds and precision, and the F1 score balances precision and recall. In the context of 18 fruit classes, achieving nearly 100 precision signifies high accuracy in classifying fruits, a recall of 0.94 indicates effective identification of most instances, and a precision-confidence of 0.76 suggests controllable precision based on confidence thresholds. The F1 score of 0.72 demonstrates a balanced performance. These values collectively imply that the model is efficient in recognizing fruit classes, making it a promising tool for fruit detection tasks, but real-world testing is crucial to validate its practical applicability. The curves of YOLOv8s are depicted in Fig. 4.

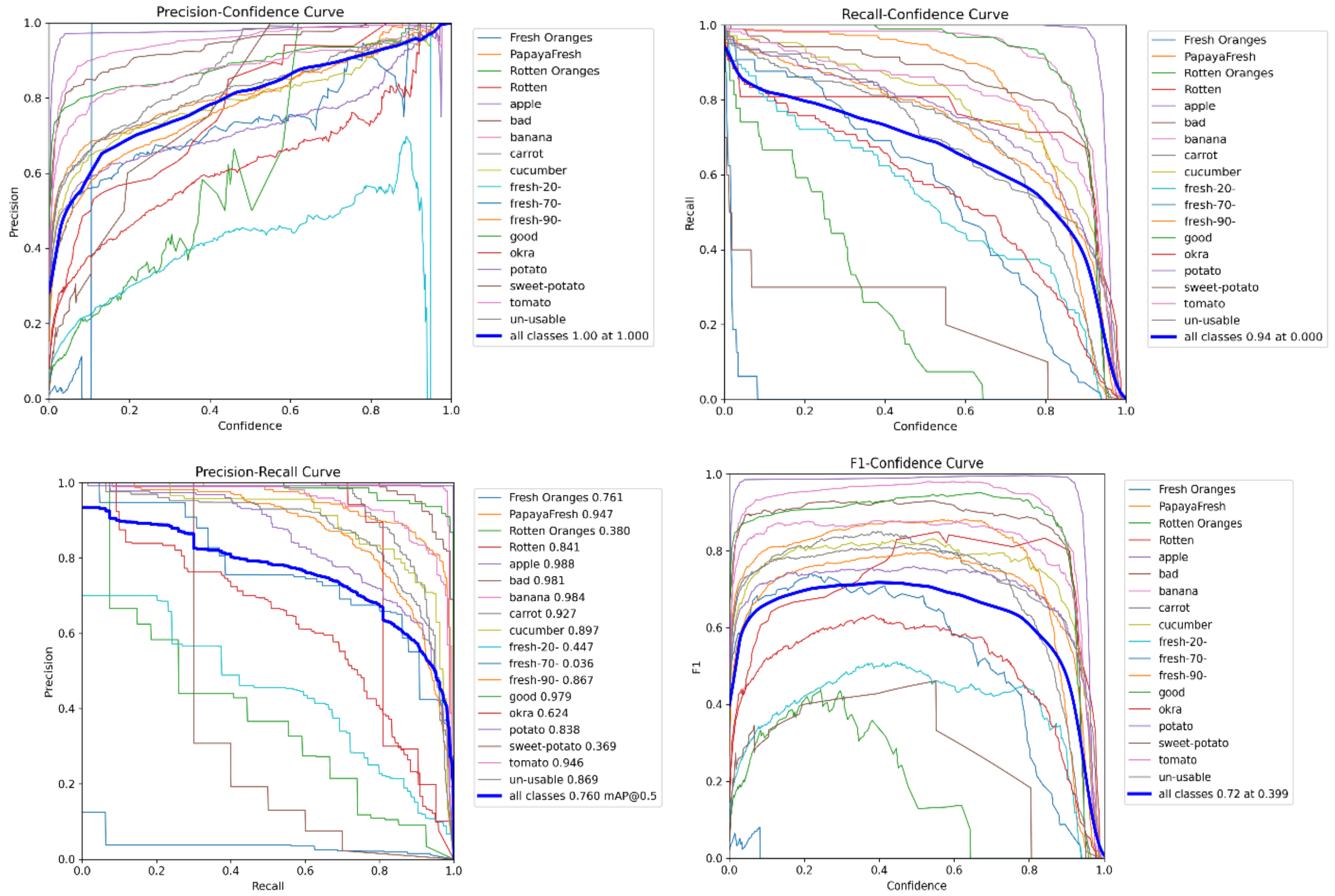


Fig. 4. The curves of YOLOv8s.

B. Model Evaluation for YOLOv8n

In the background of 18 fruit classes, achieving nearly 99 precision signifies high accuracy in classifying fruits, a recall of 0.95 indicates effective identification of most instances, and a precision-confidence of 0.73 suggests controllable precision based on confidence thresholds. Although the F1 score of 0.69 indicates a slightly lower balance between precision and recall, these values collectively indicate that the model is quite effective in recognizing fruit classes, making it a promising tool for fruit detection tasks. Real-world testing and fine-tuning may further enhance its performance for practical applications. Fig. 5 shows the curves of YOLOv8n.

C. Model Evaluation for YOLOv8l

Achieving nearly 100 precision indicates highly accurate classification of fruits, a recall of 0.96 demonstrates effective identification of most fruit instances, and a precision-confidence of 0.76 suggests controllable precision, considering confidence levels. The F1 score of 0.72 showcases a reasonable trade-off between precision and recall. These values collectively affirm that the model is highly efficient in recognizing the 18 fruit classes, making it a robust and reliable tool for fruit detection tasks, though further evaluation in real-world scenarios is advisable to confirm its practical effectiveness. The curves of YOLO8vl are depicted in Fig. 6.

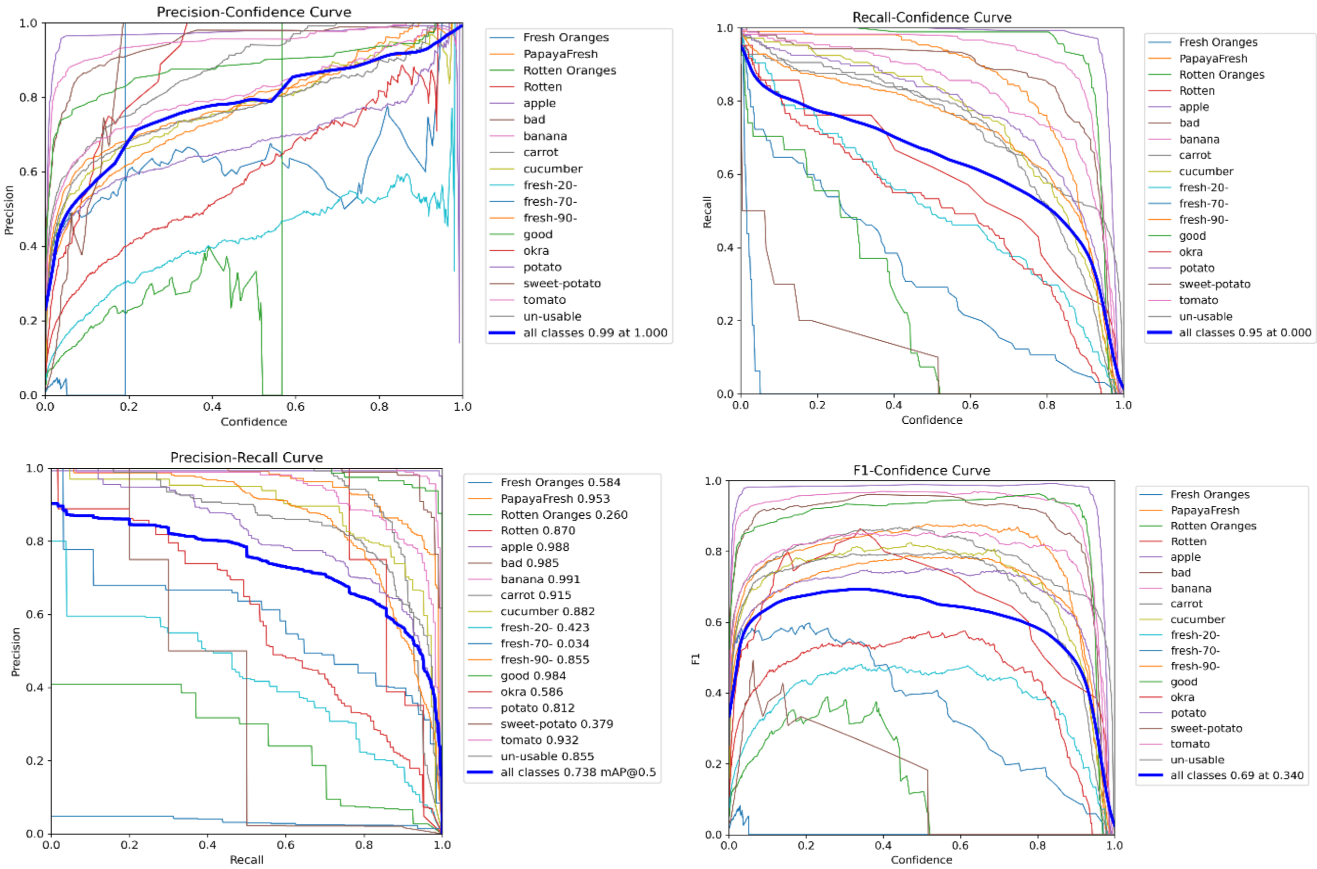
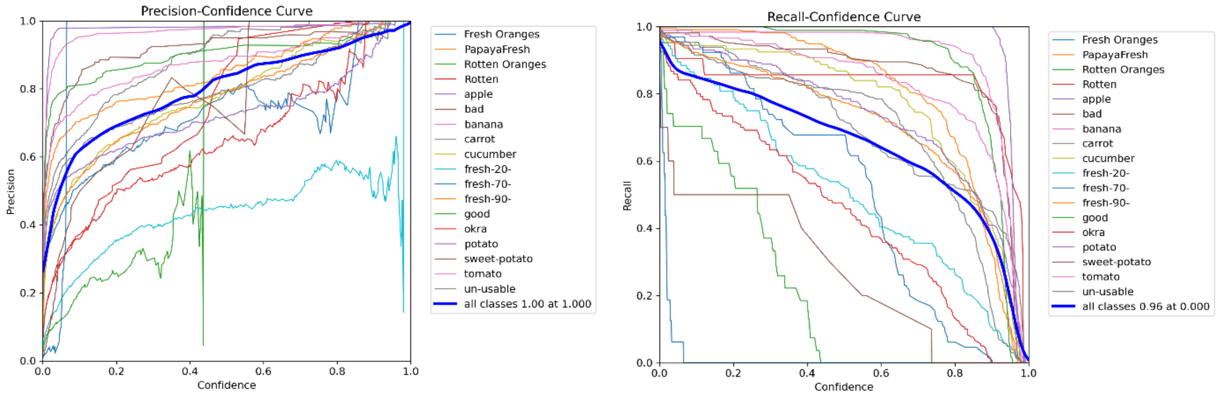


Fig. 5. The curves of YOLOv8n.



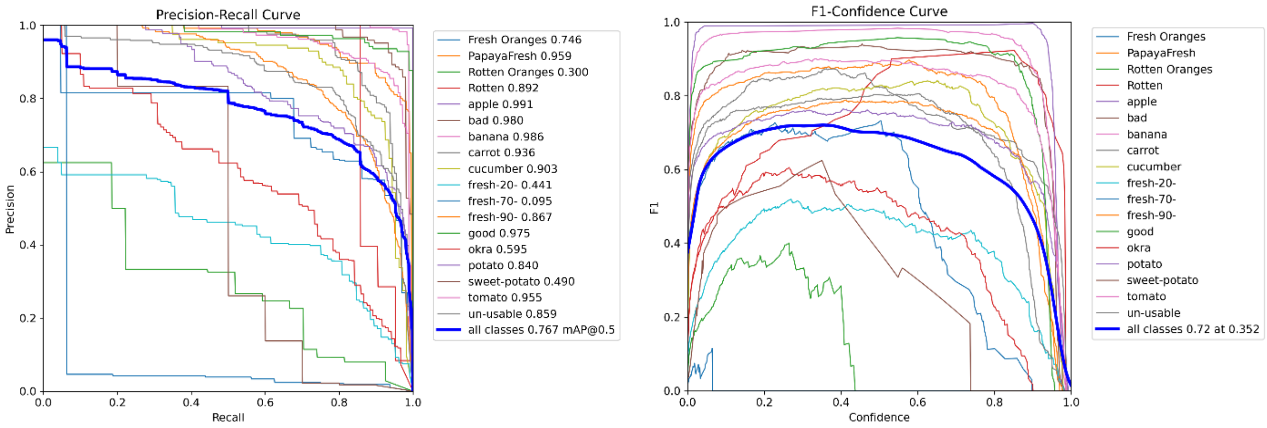


Fig. 6. The curves of YOLO8vl.

D. Model Evaluation for YOLOv8x

Achieving nearly 100 precision signifies highly accurate classification of fruits, a recall of 0.95 indicates effective recognition of the majority of fruit instances and a precision-confidence of 0.72 implies controllable precision at different confidence levels. The F1 score of 0.76 demonstrates a good overall balance between precision and recall. These values collectively suggest that the YOLOv8vx model is effective in recognizing the 18 fruit classes, making it a robust and reliable tool for fruit detection tasks with the potential for real-world applications. The curves of YOLOv8x are depicted in Fig. 7.

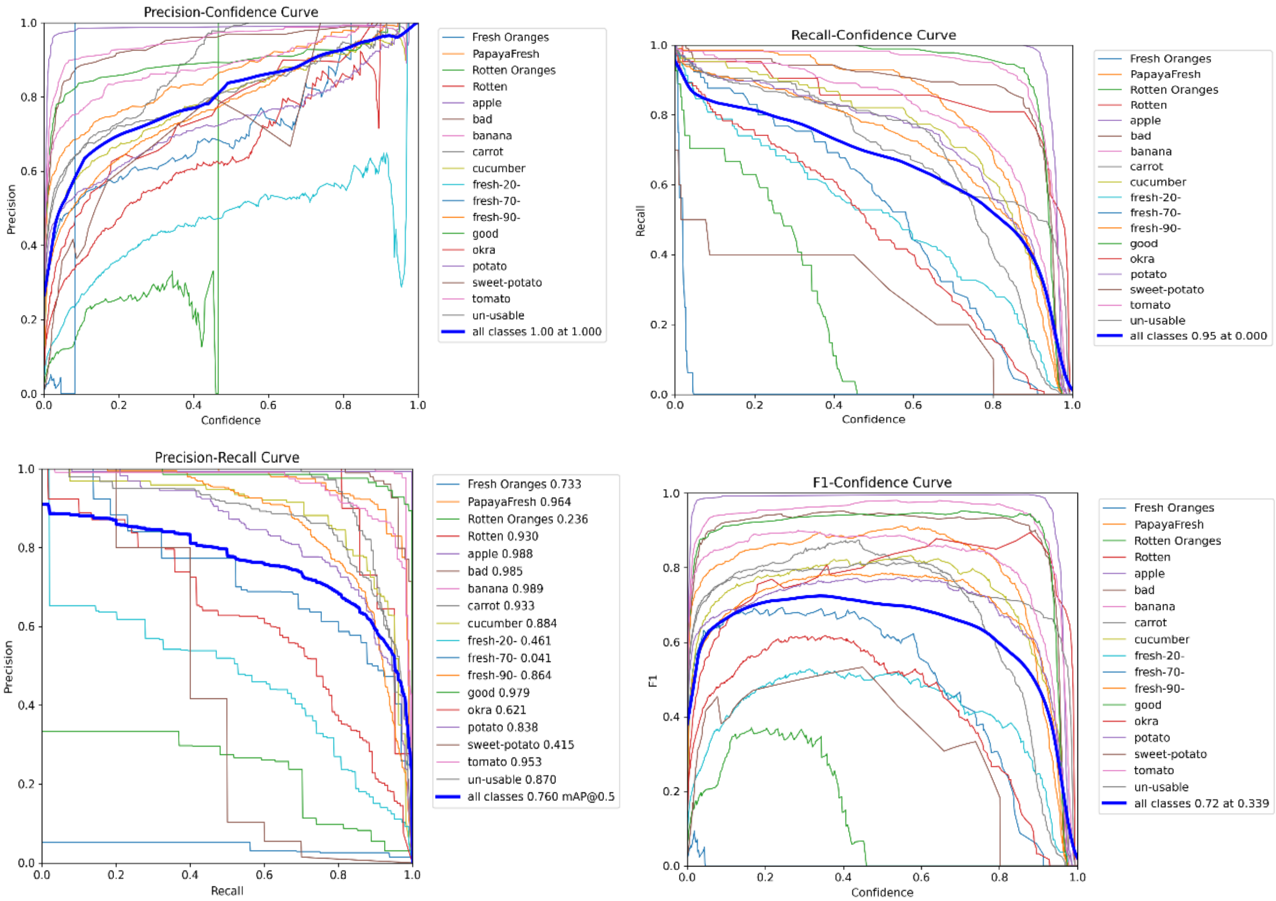


Fig. 7. The curves of YOLOv8x.

V. Results and Discussion

In the comprehensive series of experiments, rigorously assessed multiple YOLOv8 models to identify the most accurate and effective one for the specific task. The study collected performance metrics across all classes, including precision, recall rate, mean Average Precision (mAP) at an IoU threshold of 0.5, and F1 score, aiming to achieve the utmost accuracy and effectiveness in the model selection.

As discussed earlier, extensive literature supports the effectiveness of YOLO-based models in achieving high accuracy while maintaining real-time processing capabilities, making them particularly suitable for various applications. The simplicity and efficiency of the YOLO architecture have positioned it as a benchmark in the field of object detection.

In this study, the choice of Yolov8 as the foundation for the proposed method is justified by the extensive experiments conducted and the comprehensive comparison of various versions of Yolov8-based models. By presenting a detailed evaluation and comparison of different model configurations, this study aims to showcase the superiority of Yolov8 in the context of fruit detection. The experimental results contribute empirical evidence to the existing literature, reinforcing the claim that Yolov8 stands out as an effective and reliable object detection algorithm, especially when applied to the specific challenges posed by fruit detection in agriculture.

Upon analyzing the results, it is evident that the YOLOv8s, YOLOv8l, and YOLOv8x models consistently outperform the YOLOv8n model across various metrics. Notably, all three of these models achieved a perfect precision score of 100%, indicating their exceptional ability to make correct positive predictions. Furthermore, the YOLOv8l and YOLOv8x models demonstrated superior recall rates of 0.96% and 0.95%, respectively, highlighting their capacity to identify most of the actual positive instances. Additionally, these models maintained a robust mAP@0.5 rate of 0.76% and an impressive F1 score of 0.72%, signifying a balanced trade-off between precision and recall.

Considering the balance between precision and recall, the YOLOv8l and YOLOv8x models emerge as the top performers in the evaluations. Their remarkable precision, recall, and F1 score collectively demonstrate their superiority in accurately detecting and classifying objects, making them the preferred choices for the task. Based on these extensive experiments, it has successfully achieved an accurate and effective model tailored to the specific requirements. The comparison table between all versions of YOLO8 is depicted in Table I.

TABLE I. The Comparison Table between All Versions of YOLO8

Model version	F1 Score	precision	Recall	mAP @0.5
YOLOv8s	0.72	100	0.94	0.76
YOLOv8n	0.69	99	0.95	0.73
YOLOv8l	0.72	100	0.96	0.76
YOLOv8x	0.72	100	0.95	0.76

VI. Conclusion

Fruit detection holds paramount significance in the agricultural sector, aiding in the optimization of harvesting schedules, minimizing waste, and ensuring crop quality. Numerous methods have been explored in the literature to address this critical task. Among these, deep learning-based approaches have emerged as frontrunners, consistently delivering accurate results. However, a prevailing research challenge in deep learning-based fruit detection pertains to meeting the stringent accuracy rate requirements necessitated by agricultural applications. This study proposed a deep learning model based on the YOLOv8 architecture to address this challenge. Leveraging a custom dataset, it meticulously conducts model training, validation, and testing. The experimental results and performance evaluations demonstrate the efficacy of our proposed method, showcasing its ability to achieve high levels of accuracy, thus promising substantial advancements in fruit detection within the agricultural domain. Two notable limitations in the realm of fruit detection are computational resource intensity and model generalization. Firstly, deep learning-based fruit detection models often require substantial computational resources for training and inference, which may not be readily available in resource-constrained agricultural environments. Secondly, achieving model generalization across different fruit varieties, lighting conditions, and backgrounds remains a challenge, as models trained on one dataset may struggle to adapt to diverse real-world scenarios. In light of these limitations, future research could focus on addressing these challenges. Firstly, the development of more computationally efficient deep learning architectures tailored for fruit detection could help alleviate resource constraints. Secondly, exploring techniques such as domain adaptation and transfer learning to enhance model generalization across varying conditions and fruit types could lead to more robust and versatile fruit detection systems. These efforts could significantly enhance the applicability and effectiveness of fruit detection technology in agriculture.

References

[1] C. C. Ukwuoma, Q. Zhiguang, M. B. Bin Heyat, L. Ali, Z. Almaspoor, and H. N. Monday, “Recent advancements in fruit detection and classification using deep learning techniques,” Math Probl Eng, vol. 2022, pp. 1–29, 2022.

[2] F. Gao et al., “A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard,” Comput Electron Agric, vol. 197, p. 107000, 2022.

[3] W. Zhang et al., “Deep-learning-based in-field citrus fruit detection and tracking,” Hortic Res, vol. 9, 2022.

[4] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, “Deep learning–Method overview and review of use for fruit detection and yield estimation,” Comput Electron Agric, vol. 162, pp. 219–234, 2019.

[5] H. Kang and C. Chen, “Fast implementation of real-time fruit detection in apple orchards using deep learning,” Comput Electron Agric, vol. 168, p. 105108, 2020.

[6] W. Zhang, K. Chen, J. Wang, Y. Shi, and W. Guo, “Easy domain adaptation method for filling the species gap in deep learning-based fruit detection,” Hortic Res, vol. 8, 2021.

[7] M. Afonso et al., “Tomato fruit detection and counting in greenhouses using deep learning,” Front Plant Sci, vol. 11, p. 571299, 2020.

[8] K. Bresilla, G. D. Perulli, A. Boini, B. Morandi, L. Corelli Grappadelli, and L. Manfrini, “Single-shot convolution neural networks for real-time fruit detection within the tree,” Front Plant Sci, vol. 10, p. 611, 2019.

[9] L. Fu, F. Gao, J. Wu, R. Li, M. Karkee, and Q. Zhang, “Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review,” Comput Electron Agric, vol. 177, p. 105687, 2020.

[10] M. C. Ang, E. Sundararajan, K. W. Ng, A. Aghamohammadi, and T. L. Lim, “Investigation of Threading Building Blocks Framework on Real Time Visual Object Tracking Algorithm,” Applied Mechanics and Materials, vol. 666, pp. 240–244, 2014.

[11] A. Aghamohammadi, M. C. Ang, E. A. Sundararajan, N. K. Weng, M. Mogharrebi, and S. Y. Banihashem, “A parallel spatiotemporal saliency and discriminative online learning method for visual target tracking in aerial videos,” PLoS One, vol. 13, no. 2, p. e0192246, 2018.

[12] X. Wang et al., “Lightweight SM-YOLOv5 tomato fruit detection algorithm for plant factory,” Sensors, vol. 23, no. 6, p. 3336, 2023.

[13] T.-H. Liu et al., “Pineapple (Ananas comosus) fruit detection and localization in natural environment based on binocular stereo vision and improved YOLOv3 model,” Precis Agric, vol. 24, no. 1, pp. 139–160, 2023.

[14] R. Gai, N. Chen, and H. Yuan, “A detection algorithm for cherry fruits based on the improved YOLO-v4 model,” Neural Comput Appl, vol. 35, no. 19, pp. 13895–13906, 2023.

[15] J. Zhou, Y. Zhang, and J. Wang, “A dragon fruit picking detection method based on YOLOv7 and PSP-Ellipse,” Sensors, vol. 23, no. 8, p. 3803, 2023.

[16] C. B. MacEachern, T. J. Esau, A. W. Schumann, P. J. Hennessy, and Q. U. Zaman, “Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks,” Smart Agricultural Technology, vol. 3, p. 100099, 2023.

[17] A. Cardellicchio et al., “Detection of tomato plant phenotyping traits using YOLOv5-based single stage detectors,” Comput Electron Agric, vol. 207, p. 107757, 2023.

[18] E. Wei and S. Ren, “Drowning Detection based on YOLOv8 improved by GP-GAN Augmentation,” 2023.

Optimizing Mobile Ad Hoc Network Routing using Biomimicry Buzz and a Hybrid Forest Boost Regression - ANNs

D. Dhinakaran[1*], S.M. Udhaya Sankar[2], S. Edwin Raja[3], J. Jeno Jasmine[4]

Department of Computer Science and Engineering,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India[1, 3]
Department of CSE (Cyber Security), R.M.K College of Engineering and Technology, Chennai, India[2]
Department of Computer Science and Engineering, R.M.K. Engineering College, Tamil Nadu, India[4]

Abstract*—A mobile ad hoc network (MANET) is a network of moving nodes that can interact with one another without the aid of a centrally located infrastructure. In MANETs, every node acts as a router and as a host, generating and consuming data. However, due to the mobility of nodes and the absence of centralized control, the routing process in MANETs is challenging. Therefore, routing protocols in MANETs are required to be efficient, scalable, and adaptable to the dynamic topology changes of the network. This paper proposes an optimized route selection approach for MANETs via the biomimicry buzz algorithm with the Bellman-Ford-Dijkstra algorithm to improve the effectiveness and accuracy of the routing process. By integrating these behaviors into the algorithm, the approach can select the shortest path in a network, leading to an optimal routing solution. Furthermore, the paper explores the use of Forest Boost Regression (FR), a novel machine learning algorithm, to predict energy consumption in MANETs. Utilizing this will help the network run more efficiently and last longer. Additionally, the paper discusses the use of Artificial Neural Networks (ANNs) to forecast link failure in MANET s, thereby increasing network performance and dependability. The proposed work presents the experimental evaluation by using Ns-3 as the simulation tool. The experimental results indicate a variation in packet delivery ratio from 97% to 90%, an average end-to-end delay of approximately 19 ms, an increase in node speed energy consumption from 60 to 87 joules, and a simulation time energy consumption of 89 joules over 60 seconds. These results provide insights into the performance and efficiency of the proposed strategy in the context of MANETs.

Keywords—MANET; routing protocols; optimized route selection; regression; machine learning; Artificial Neural Networks

I. INTRODUCTION

The nodes in MANETs, which are wireless networks with no infrastructure, can interact with one another directly or indirectly [1]. MANETs are widely used in various applications such as military, disaster response, and emergencies, where infrastructure is not available. Due to the lack of infrastructure, in MANETs, the nodes must collaborate to provide network communication [2-4]. Therefore, because of the constantly changing topology, scarce resources, and frequently occurring linkage failures, routing in MANETs is a difficult task. Proactive, reactive, as well as hybrid routing algorithms may all be categorized in MANETs. Proactive procedures keep all nodes' routing information current, even without traffic [5]. Reactive protocols establish routes on demand only when needed. Proactive as well as reactive protocols' benefits are combined in hybrid protocols [6-8]. Although routing protocols have advanced, routing in MANETs remains challenging because of the network's continually changing character. One of the significant challenges in MANETs is predicting link failures. Link failures occur due to various reasons, such as node mobility, interference, and limited battery power [9]. A routing protocol that can predict link failures and react accordingly can significantly improve the performance of MANETs [10-12]. In addition, energy utilization is another crucial factor in MANETs since nodes have limited battery power.

In a typical MANET scenario, a diverse set of wireless devices, often with varying mobility patterns, come together to form a self-organizing and infrastructure-less network as shown in Fig. 1. These devices could include smartphones, laptops, IoT sensors, or even military communication devices. MANETs are often deployed in environments where traditional fixed infrastructure is unavailable or impractical, such as disaster-stricken areas, military operations, or highly dynamic urban settings [13-15]. Nodes in a MANET are both end-users and routers, capable of transmitting and forwarding data packets to facilitate communication among themselves [16]. The unique feature of MANETs is their ad hoc nature, and the network topology continually changes due to node mobility [17]. This dynamic topology, along with the absence of centralized control, presents routing challenges, where finding efficient and reliable communication paths are paramount [18-20]. MANETs offer a flexible and resilient solution for communication in dynamic and often challenging environments, but effective routing, energy management, and reliability remain critical considerations to ensure their successful operation.

In our work, the primary objectives of optimizing MANET routing are to enhance adaptability, reduce latency, and improve overall network performance. Specifically, our proposed approach aims to achieve the following goals:

- Enhance the routing protocol's adaptability to the dynamic topology changes inherent in MANETs.

- Enable efficient decision-making in routing, considering factors like node mobility, link stability, and energy levels.
- Mitigate issues related to link breakage and broadcast storms by introducing a mobility-aware routing algorithm.
- Minimize packet delivery delay by optimizing the routing process.
- Optimize route selection to achieve more efficient and reliable communication between nodes.

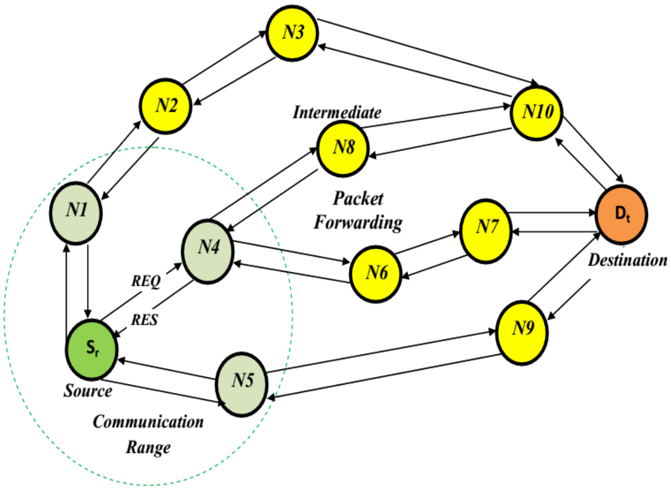


Fig. 1. Classic MANET system – scenario.

By addressing these goals, our work aims to contribute to the advancement of MANET routing protocols, making them more adaptive, responsive, and efficient in dynamic and challenging environments. Our work introduces an optimized route selection approach that addresses the routing challenges in MANETs. We combine the biomimicry buzz algorithm with the well-established Bellman-Ford-Dijkstra algorithm to improve the efficiency and accuracy of route selection. By incorporating these behaviors into our approach, we aim to identify the shortest paths in the network, leading to optimal routing solutions. Moreover, we recognize the importance of energy management in MANETs, as it directly impacts network sustainability and performance. To this end, our research explores the application of Forest Boost Regression (FR), a novel machine learning algorithm, for predicting energy consumption in MANETs. Accurate energy consumption predictions can enable more efficient resource allocation and contribute to prolonged network operation.

Furthermore, network reliability is another crucial aspect in MANETs. Link failures can disrupt communication and hinder the network's effectiveness [21-23]. To address this, our work investigates the use of Artificial Neural Networks (ANNs) to predict link failures, thereby enhancing network performance and reliability. Our study employs Ns-3 as the simulation tool for experimental evaluation. We assess the proposed approach using various metrics taking into account factors. Through this research, we aim to provide valuable insights into the performance and efficiency of our proposed strategy, offering solutions to the routing, energy consumption, and reliability challenges that MANETs face in their dynamic and infrastructure-less environment.

In the subsequent sections, we delve into an extensive literature survey in Section II, where we analyze existing works related to Mobile Ad Hoc Networks (MANETs) and routing protocols. Section III unfolds our proposed work, presenting the Biomimicry Buzz Algorithm, the Optimized route selection using the Bellman-Ford-Dijkstra algorithm, as well as our predictive models for Energy Consumption and Link Failure. These sections provide a comprehensive insight into the novel contributions and methodologies we propose. Following that, Section IV details the performance analysis, where we evaluate our approach through rigorous experimentation. Finally, Section V concludes the paper, summarizing key findings, limitations, and outlining potential avenues for future research.

II. Literature Survey

In the evolving landscape of Mobile Ad Hoc Networks (MANETs), persistent challenges include routing inefficiencies, prompting ongoing research into novel solutions. Recent efforts have yielded innovative routing protocols, aiming to address the shortcomings of conventional protocols like AODV and DSR, which often struggle in dynamic network topologies. Moreover, energy management techniques have traditionally relied on rule-based approaches, but emerging trends embrace machine learning algorithms for energy prediction and optimization. Despite these advancements, limitations such as scalability and security persist, warranting further investigation into comprehensive and robust solutions to bolster the reliability, efficiency, and sustainability of MANETs in dynamic and infrastructure-less environments.

In the study by Kai et al. [24], the focus lies on identifying the optimal algorithm path for quality routing in Ad-hoc networks. They proposed Hopfield neural network model that addresses the minimum cost problem with time delay. By carefully choosing these values, the enhanced path algorithm establishes the relationship between energy function parameters and shows that the network's possible solution falls within the heading of progressive stability. The calculations show that the answer is independent of the starting value and constantly produces the world's best solution. An adaptive routing protocol with bio-inspired design is introduced by Shah et al. in [25]. The AOMDV-FG technique maximizes a number of paths derived from the AOMDV mechanism, choosing the best path based on the highest fitness value. By comparing it against the AOMDV-TA and EHO-AOMDV protocols using important metrics, the authors evaluate the performance of their model.

The Trust and ANT Based Routing (TABR) technique is suggested for MANETs by Sridhar et al. [26]. An ant-based routing algorithm and trust values are combined by TABR to find reliable, trusted, and optimized routes in the network. TABR seeks to improve routing efficiency by combining the benefits of trust mechanisms and ant-based routing. Alsaqour et al. [27] provide the genetic algorithm-based location-aided routing algorithm to increase the effectiveness of MANET routing protocols. In order to improve delivery behavior, GALAR uses genetic optimization and adaptive updates of node position information. With little network overhead, it delivers a high packet delivery ratio of over 99%. Jeena Jacob et al. [28] method of optimizing performance using the proper tools, building a wireless environment-specific model, and enhancing routing through the careful selection of performance indicators are the three main phases they suggest. In order to enable synchronous and decentralized routing decisions, they make use of the Artificial Bee Colony Optimization method, which displays basic bee agent behaviors. The benefits of their work are found in the ABC algorithm's beneficial effects on wireless network connectivity. Their suggested method improves the network's overall effectiveness and performance by utilizing its evaluative qualities.

The Ant Colony Optimization (ACO) technique is used by Dorathy et al. [29] to handle network issues with routing as well as protection in networks that are both wired and wireless. They concentrate on finding the network's shortest, most efficient path connecting the point of origin and the destination. Their routing strategy attempts to increase the overall lifespan of the entire network by taking into account variables such node reserve energy, least residual power of the path, node distance, trip time, and hop count. The benefits of their work include taking into account a variety of variables to choose the best path, and increasing energy efficiency. The intelligent Whale Optimization Algorithm (p-WOA), presented by Husnain et al. [30], is a cluster-based, bio-inspired algorithm for routing in vehicular communication. By including factors like communication span, velocity, and path along the highway in the fitness function, the p-WOA algorithm lowers randomness and enhances cluster head (CH) selection. Their research shows that the p-WOA technique outperforms conventional approaches like the Ant Lion Optimizer along with Grey Wolf Optimization in terms of obtaining the ideal number of cluster heads. EHO-AOMDV is a routing algorithm that Sarhan et al. [31] introduce with the goal of minimizing the total energy. To lessen the likelihood of path malfunction and the number of dead nodes brought on by heavy data loads, nodes are divided into two groups. The use of energy-aware classification and updating techniques to increase the effectiveness of energy is one of the benefits of their work. For VANETs, Muhammad et al. [32] suggest a grey wolf optimization-based clustering approach. To produce effective clusters. An optimized number of clusters is produced as a result of earlier convergence caused by the grey wolf nature's linearly declining component. Their work has advantages in that it incorporates behaviors that are inspired by nature and make it possible to cluster data in VANETs effectively.

Hybrid ant and bee colony optimization algorithm, a method for picking the cluster head in MANETs, is presented by Janakiraman et al. [33]. The drawbacks of ACO in addition to ABC are addressed by this method by integrating them in a complementary way. Their method intends to avoid stagnation in the intensification process of ACO and address delayed convergence in the spectator bee phase of ABC by using employee bee agents for dividing the method of extraction into two levels. The enhanced choice of cluster head procedure is where their work excels. By streamlining complex algorithms like the Bat Optimization Algorithm, Particle Swarm Optimization, and Ant Colony Algorithm for distance optimization, Charan et al. [34] concentrate on improving routing algorithms in MANETs. They suggest a protocol called Bat Optimized Link State Routing. BOLSR tries to identify the best path by exchanging precise messages by fusing the OLSR structure and the Bat Algorithm. This results in the BOLSR protocol, which calculates the best route between the nodes based on their energy characteristics. A routing system for MANETs is proposed by Junnarkar et al. [35] and is based on the Ant Colony based optimization algorithm. The ACO-based method improves Quality of Service (QoS) effectiveness by using the nodes' present location and load factors as routing metrics. RSSI data are used in their proposed QoS Mobility Aware ACO Routing Protocols to calculate the separation between mobile nodes.

Based on the literature survey, it can be concluded that various bio-inspired and machine learning-based routing algorithms have been proposed for MANETs to enhance the routing process's effectiveness, accuracy, and reliability. The ACO, PSO, bee colony optimization, and bat algorithm have been widely used for optimizing the routing path. In comparison to the previous works, the proposed methodology offers several advantages and advancements. Firstly, it introduces an optimized route selection approach by integrating the biomimicry buzz algorithm with the Bellman-Ford-Dijkstra algorithm. This integration improves the effectiveness and accuracy of the routing process, enabling the selection of the shortest path and leading to an optimal routing solution. Additionally, the proposed methodology explores the use of ForestBoost Regression (FR) for energy consumption prediction and Artificial Neural Networks (ANNs) for link failure prediction. By utilizing machine learning techniques, the methodology enhances network efficiency and reliability. Compared to existing methods like TABR, GALAR, and others, the proposed methodology stands out by offering a comprehensive approach that addresses the limitations and challenges in MANET routing. It leverages integrated algorithms, prediction techniques, and advanced optimization methods to achieve superior performance, reliability, and energy efficiency.

III. Proposed Work

A. Motivation

The motivation for our proposed work stems from the pressing need to enhance the performance and sustainability of MANETs in dynamic and infrastructure-less settings. MANETs are increasingly deployed in scenarios where traditional network infrastructure is absent or impractical, such as disaster response, military operations, and highly mobile urban environments. The inherent mobility of nodes and the lack of a centralized control infrastructure in MANETs present unique challenges, particularly in the realms of routing efficiency, energy consumption, and network reliability. Inefficient routing protocols can lead to high overhead and significant delays, energy depletion can curtail network longevity, and unpredictable link failures can disrupt communication. To address these issues, our work aims to introduce innovative solutions, combining the Bellman-Ford-Dijkstra algorithm with biomimicry-inspired routing, employing machine learning for energy consumption prediction, and using Artificial Neural Networks for link failure prediction. The ultimate motivation is to contribute to the development of comprehensive and data-driven strategies that improve the performance, sustainability, and reliability of MANETs, thus extending their utility in critical and challenging environments. By addressing these issues, our work aligns with the broader objective of advancing the state-of-the-art in MANET research and facilitating more effective communication in dynamic and infrastructure-less networks.

The main point of this article is summed up in the facts mentioned below:

1) Bio-inspired navigation module: This module includes the algorithms for simulating the natural behaviors of honeybee waggle dance and bat echolocation, named biomimicry buzz algorithm. It receives input data such as the location and distance of the nodes and outputs the direction and distance to the destination node.

2) Bellman-ford-dijkstra algorithm module: This module includes the algorithms for discovering the fastest path among two nodes in the network. It receives input data such as the network topology and the output of the honeybee waggle dance and bat echolocation behavior module and outputs the optimized routing solution.

3) ForestBoost regression energy consumption prediction module: This module includes the algorithms for predicting the energy consumption of nodes in the network using the ForestBoost Regression machine learning algorithm. It receives input data such as historical data and network topology and outputs the predicted energy consumption.

4) Artificial Neural Networks link failure prediction module: This module includes the algorithms for predicting link failures in real-time using Artificial Neural Networks. It receives input data such as historical data and network topology and outputs the predicted link failure probability.

The proposed Bio-inspired navigation approach combines the natural behaviors of honeybee waggle dance and bat echolocation with the Bellman-Ford-Dijkstra algorithm to select the shortest path in a network, leading to an optimal routing solution. The approach considers the source node, destination node, higher value node, and intermediate nodes to determine the optimal route for packet transmission, as shown in Fig. 2. The use of a Bio-inspired navigation approach provides an efficient and reliable mechanism for route selection and to avoid congested or noisy paths in MANETs.

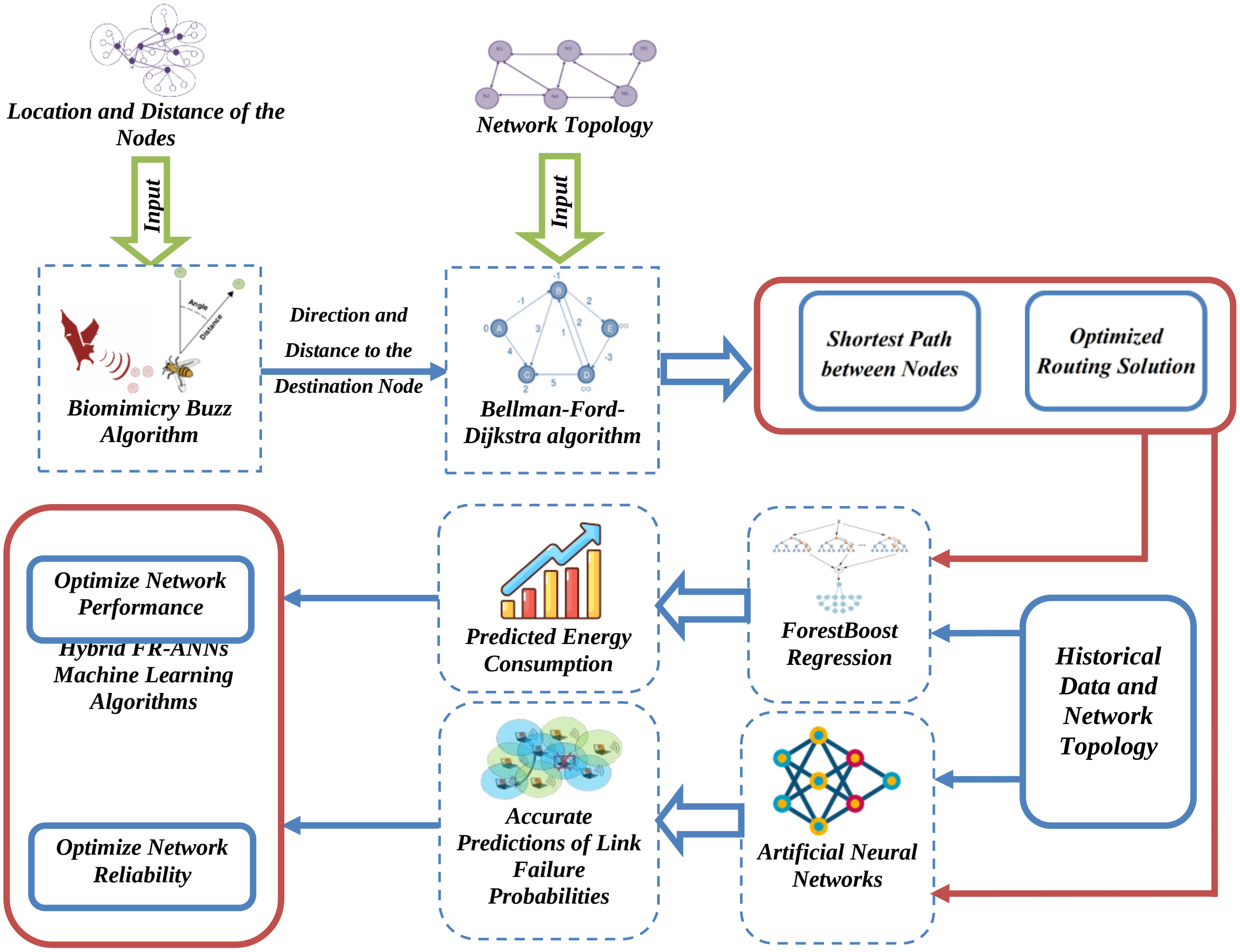

Location and Distance of the Nodes
Network Topology
Input
Input
Angle
Distance
Biomimicry Buzz Algorithm
Direction and Distance to the Destination Node
Bellman-Ford-Dijkstra algorithm
Shortest Path between Nodes
Optimized Routing Solution
Optimize Network Performance
Hybrid FR-ANNs Machine Learning Algorithms
Optimize Network Reliability
Predicted Energy Consumption
ForestBoost Regression
Historical Data and Network Topology
Accurate Predictions of Link Failure Probabilities
Artificial Neural Networks

Fig. 2. Complete architecture of proposed model.

In addition, the paper explores using FR-ANN, a novel machine learning algorithm, to predict energy consumption and link failure in MANETs. FR-ANN can be used to optimize the network's energy usage, prolong the network's lifespan, and enhance the network's performance and reliability by predicting potential link failures before they occur. By predicting energy consumption and potential link failures, the network can use energy-efficient routes, resulting in longer battery life for nodes, preventing data loss, and improving network availability.

B. Biomimicry Buzz Algorithm

The honeybee waggle dance is a behavior used by honeybees to communicate the location of food sources to other bees in the hive. A bee that has found food returns to the hive and performs a dance, where the angle, along with the duration of the dance, is a sign of the direction plus distance of the food source, correspondingly. Other bees in the hive use this information to locate the food source. Bat echolocation is a behavior bats use to navigate and find food in the dark. Bats emit high-frequency sounds, which bounce off objects in their environment and create echoes. They listen to the echoes to determine the distance and location of things. To integrate these two behaviors for route selection, we could use the honeybee waggle dance to communicate the establishment of a destination, while bat echolocation could be used to navigate around obstacles and avoid collisions.

The echolocation method used by bats involves transmitting sound waves and detecting their reflection to locate prey based on the characteristics of the echo. Similarly, a device discovery algorithm uses a similar concept to find proximal devices within a specific search space. The discoverer device sends a signal or relies on the base station to determine the location of nearby devices. Mathematically, this technique can be expressed as follows: Let L denotes the position of a bat at time tm and s denotes the speed, which represents the rate of change in status. Consequently, the bat's position-refreshing method is determined in Eq. (1)

$$L_j(tm) = L_j(tm + 1) + s_j(tm) \quad (1)$$

Bats change their trajectory and pace depending on how far they are from their prey's location, as evidenced by their characteristics, which are represented in Eq. (2)

$$S_j(tm) = S_j(tm + 1) + (L_j(tm) – L') * fn_j \quad (2)$$

L′ deals with the location of the prey, whereas fn deals with the periodicity of natural waves. The following equation can be used to simulate the irrational behavior of bats, such as migration far from their identified prey via a degree of divergence in the range of loudness by employing Eq. (3)

$$L_{new} = L_{old} + \varepsilon * Sig^{tm} \quad (3)$$

The loudness evaluation algorithm indicates a uniform distribution, and Sigtm stands for signal strength by using the Eq. (4)

$$Sig_j(tm+ 1) = \tilde{e} * Sig_j(tm) \quad (4)$$

where, Sigj(tm+ 1) → 0, where tm → ∞, and ẽ the empirical value. Bats can determine how far they are from their target by varying the frequency associated with emanation. The emanation change occurs at the following frequency by using the Eq. (5)

$$c_j(tm + 1) = c_j(0)[1 - f^{(-d_*tm)}] \quad (5)$$

while $c_i(tm) \rightarrow c_i(0)$ as tm → ∞, as well as d is the empirical constant.

Process Flow:

1) *A* node in the network needs to communicate with another node at a specific location.

2) *The* node sends a request to the network for the location of the destination.

3) *Another* node in the network, which has knowledge of the location of the destination, responds by performing a "waggle dance" to indicate the direction and distance of the goal.

4) *The* requesting node uses this information to navigate toward the destination using bat echolocation to detect obstacles and avoid collisions.

5) *As* the node gets closer to the destination, it can continue using bat echolocation to refine its navigation and avoid any obstacles.

Combining the honeybee waggle dance and bat echolocation can create a more robust and adaptive approach for route selection in a MANET. The honeybee waggle dance provides information about the location of the destination, while bat echolocation enables nodes to navigate around obstacles and avoid collisions in real time. This approach could be beneficial in situations where nodes are mobile, and the network topology is constantly changing.

C. Optimized Route Selection - Bellman-Ford-Dijkstra Algorithm

To determine the fastest path through a network of paths, the Bellman-Ford-Dijkstra procedure incorporates the Bellman-Ford algorithm and Dijkstra's algorithm. The network's negative cycles are initially detected along with removed using the Bellman-Ford algorithm, followed by the shortest route among the two nodes is determined using Dijkstra's algorithm. The algorithm effectively discoveries the shortest path in various network topological structures by integrating the advantages of each algorithm. For example, to incorporate honeybee waggle dance and bat echolocation into the Bellman-Ford-Dijkstra algorithm, we can use these behaviors to provide additional information about the network and guide the search toward the shortest path.

Use honeybee waggle dance to estimate the direction and distance of the destination node. Honeybees use this behavior to communicate the location of food sources to other bees in the hive. We can use a similar approach to estimate the location of the destination node in the network. Use bat echolocation to identify obstacles and congested areas in the network. Bats use echolocation to navigate in the dark and avoid collisions, and we can use a similar approach to help nodes navigate around obstacles and avoid congested areas. Utilize the Bellman-Ford method to find negative network cycles and eliminate them. Negative cycles are loops in the network with a negative weight, and they can cause the algorithm to enter an infinite loop. By detecting and eliminating negative cycles, we can ensure that the algorithm converges to a valid shortest path.

To determine the shortest route between two locations, use Dijkstra's algorithm. In most instances, the Dijkstra algorithm is better than the Bellman-Ford algorithm at choosing the shortest path. Using honeybee waggle dance and bat echolocation to guide the search toward the destination node and avoid obstacles, we can further optimize the search process and reduce the search space. Add the shortest route to the routing database among every network combination of points. Then, every node consults the routing database to identify the subsequent hop on the fastest way to the target node. By integrating honeybee waggle dance and bat echolocation into the Bellman-Ford-Dijkstra algorithm, we can create an optimized route selection approach that takes advantage of these natural behaviors to progress the excellent organization in addition to the accuracy of the routing process. Table I shows the example of a routing table with sample data, we have a network with six nodes, and the routing table demonstrates the shortest path between each and every pair of nodes. For instance, to reach node B from node A, the next hop is node C. The next hop is node E to reach node C from node D. Each node in a network may utilize the routing database to find the following route on the most convenient path to the target node. Note that this is just a sample table, and the routing table would be much larger and more complex in a real-world network. A regular update of the table was additionally required as the network's topography changed.

An optimized routing solution can provide essential input data for the ForestBoost Regression energy consumption prediction module and the Artificial Neural Networks link failure prediction module. The Bellman-Ford-Dijkstra module uses the output of the honeybee waggle dance and bat echolocation behavior module to determine the network's quickest route among two nodes, which results in an optimized routing solution. This optimized routing solution can be used as input for the ForestBoost Regression energy consumption prediction module to forecast the system's nodes' energy usage.

TABLE I. EXAMPLE OF A ROUTING TABLE WITH SAMPLE DATA

Source Node	Destination Node	Next Hop
A	B	C
A	C	F
B	A	C
B	C	D
C	A	F
C	B	A
D	A	B
D	C	E
E	A	C
E	B	A
F	A	C
F	B	D

Let S_p, αS_p, C_{pt}, C_{pr} indicate the power of the transmitted signal, the power of the amplifier, the power of the circuit at the transmitter, along with the power of the circuit at the receiver, respectively by Eq. (6) to Eq. (8).

$$S_p = \frac{(4\pi)^2 d_m^{\alpha'} M_l N_f}{G_t G_r \gamma^2} E_b R_b \tag{6}$$

$$\mathrm{C_{pt}} = \mathrm{P_{mix}} + \mathrm{P_{syn}} + \mathrm{P_{filt}} + \mathrm{P_{DAC}} \tag{7}$$

$$C_{pr} = P_{mix} + P_{syn} + P_{LNA} + P_{filr} + P_{ADC} + P_{IFA} \tag{8}$$

The amount of power used while transmitting in active mode (A_{pt}) is able to be determined by Eq. (9).

$$A_{pt} = S_p + \alpha S_p + C_{pt} = (1+\alpha)S_p + C_{pt} = \frac{\varphi}{\tau} S_p + C_{pt} \tag{9}$$

During reception (A_{pr}), the electricity used in the active phase is able to be provided by Eq. (10).

$$A_{pr} = C_{pr} \tag{10}$$

The ForestBoost Regression energy consumption prediction module utilizes historical data and network topology as input to predict the energy consumption of nodes. Using an optimized routing solution as input, the module can consider the energy consumption associated with specific routing decisions and provide more accurate predictions. Similarly, the Artificial Neural Networks link failure prediction module utilizes historical data and network topology to predict link failures in real-time. By incorporating an optimized routing solution as input, the module can consider the impact of routing decisions on the network's overall reliability and provide more accurate predictions of link failure probabilities.

Let E_e^i represent the starting energy concerning a node n_e and the amount of energy needed for transmission (E_e^{Tx}) or as reception (E_e^{Rx}) beginning at that node n_e. The specified L bits are actually by Eq. (11) and Eq. (12).

$$(E_e^{Tx}) = \mathrm{P_{tx}T_{tx}} = \frac{\varphi}{\tau}\mathrm{P_t P_{ct}} \tag{11}$$

$$(E_e^{rx}) = \mathrm{P_{rx}T_{rx}} = \mathrm{P_{cr}T_{rx}} \tag{12}$$

After transmitting L bits, the RE (E_k^{Tx}) at a node n_e is given by Eq. (13).

$$(E_e^R) = \begin{Bmatrix} E_e^i - E_e^{Tx} \\ E_e^i - E_e^{Rx} \end{Bmatrix} \tag{13}$$

When using the approach we suggest with Z hops, the total amount of energy used per iteration by the fastest route via the node's $n_l(n_{lo}, n_{l1}, n_{l2}, \ldots, n_{lz},)$ can be determined based on Eq. (14).

$$E_T = \sum_{l=0}^{Z-1} [(E_k^{Tx} + E_{l+1}^{Rx})] + \sum_{l=0}^{Z} [E_l^{idle} + E_l^{CPU} + E_l^{Bat} + E_l^{DC}] \tag{14}$$

where, $E_l^{idle} = P_{idle} T_{idle}$ at node n_l and $P_{sp} \approx 0$.

Optimized route selection - Algorithm

```
1. Let s be the source node along with t is the destination node.
2. Use the honeybee waggle dance to estimate the direction
```

and distance of the destination node. Let θ be the anticipated angles among the present point and the target node, along with let the d_s is the anticipated distance among the two nodes.

3. Use bat echolocation to identify obstacles and congested areas in the network. Let $w_t(e)$ be the edge-weight e in a network, which represents the cost of traversing that edge. If an obstacle or congested area is detected along edge e, then set $w_t(e)$ to a high value to discourage nodes from using that edge.

4. Initialize the distance estimates as well as predecessor nodes for each node in the network as follows:

$d_s[s] = 0$

$d_s[v] = \infty$ for all other nodes v in the network

p[v] = null for all nodes v in the network

5. Utilize the Bellman-Ford method to find negative network cycles and eliminate them. Iterate over all edges e in the network |V|-1 times (where |V| represents the network's total number of connections), and update the distance estimates and predecessors as follows:

if $d_s[v] > d_s[u] + w_t(e)$ then:

$d_s[v] = d_s[u] + w_t(e)$ p[v] = u

If after |V|-1 iterations, there exists an edge e with $d[v] > d[u] + w_t(e)$, subsequently a negative cycle exists in the network and terminates the algorithm.

6. Using Dijkstra's algorithm, find the quickest route among the starting node s along with target node t. Initialize a priority queue Q with all nodes in the network, ordered by their distance estimates $d_s[v]$. Set $d_s[s] = 0$ and insert s into Q. Then, whereas Q is not empty, extract the node u with the smallest distance estimate d[u] from Q and relax all its outgoing edges e as follows:

if $d_s[v] > d_s[u] + w_t(e)$ then: $d_s[v] = d_s[u] + w_t(e)$ p[v] = u

If v = t, then the shortest path has been found and the algorithm terminates.

7. If a path does not connect the source node s along with the target node t, then the algorithm terminates and returns "no path".

D. Energy Consumption Prediction

Energy consumption prediction is an essential aspect of optimizing route selection in MANETs. By predicting the energy consumption of different routes, we can choose the route that requires the least amount of energy, thereby conserving battery power and extending the life of the network's components. In the approach that integrates honeybee waggle dance and bat echolocation into the Bellman-Ford-Dijkstra algorithm, we use Random Forest Regression as a hybrid machine learning algorithm to predict the energy consumption of each route. This is done by training the algorithms on historical energy consumption data for different routes and using the trained models to predict the energy consumption of new routes. The predicted energy consumption values are then integrated into the Bellman-Ford-Dijkstra algorithm to find the route with the lowest energy consumption. This route is then selected as the optimized route for data transmission in the network.

E. Random Forest Regression

An effective machine learning approach called Random Forest Regression may be utilized to anticipate energy consumption. A collaborative algorithm for learning called Random Forest Regression builds several decision trees and integrates their forecasts to create a more precise and trustworthy approach. It is well-suited for handling high-dimensional data and can control categorical and continuous variables. Gradient Boosting Regression is another ensemble learning algorithm that builds models sequentially. It is capable of handling several kinds of data and has been shown to be effective in predicting energy consumption.

Let $T_b(i)$ denotes the predicted output of Decision Tree T_b, for sample i. To forecast at a fresh location p:

$$\text{Regression: f\^{}rf B (p)} = f_{rf}^{B}(p) = \frac{1}{\mathrm{B}} \sum_{\mathrm{b}=1}^{\mathrm{B}} \mathrm{T_b(p)} \qquad (15)$$

Classification: Let $C_b(p)$ is the class predicted of b^{th} random-forest tree.

$$\text{Then } C_{rf}^{B}(p) = majority\ vote\{C_b(p)\}_1^B \qquad (16)$$

A constant value is determined for every one of the discontinuous sections that make up the input space. There are j leaf nodes in each random forest regression tree. The $g_m(p)$ values for the Random Forest Regression tree is achieved based on Eq. (17) to Eq. (19).

$$\mathrm{gm(p)} = \sum_{\mathrm{j}=1}^{\mathrm{j}} (\mathrm{b_{jm} I}), \mathrm{p} \in \mathrm{R_{jm}} \qquad (17)$$

$$\mathrm{I(p \in R_{jm})} = \begin{Bmatrix} 1, \mathrm{p} \in \mathrm{R_{jm}}; \\ 0, other; \end{Bmatrix} \qquad (18)$$

$$\mathrm{Z(L, f(p))} = \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{L} - \mathrm{f(p)})^2 \qquad (19)$$

Step 1: Initialization of Model:

$$\mathrm{fo(p)} = \mathrm{argmin} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Z(y_i, p)} \qquad (20)$$

Step 2: R random Forest regression trees are generated iteratively, with r denoting the r^{th} tree to stay r=1 to R:

(1) J represent the j^{th} selection for j = 1 to N. Finally, the loss function's low gradient number is determined, along with the result can be utilized to measure the residual r_{jr}:

$$r_{jr} = \left[\frac{\partial Z(yj, f_{r-1}(p_j))}{\partial f_{r-1}(p_j)}\right]_{f(p)} = f_{r-1}(p) \qquad (21)$$

(2) A random Forest regression tree gm(p) is produced on behalf of the enduring created in the preceding phase. The amount of steps for gradient decline is subsequently established by dividing the input space regarding the r-tree towards J distinct regions, designated as D_{1r}, D_{2r}, :::, and D_{jr}:

$$p_{r,} = \mathrm{argmin} \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{Z}(\mathrm{y_j}, f_{r-1}(p_j) + pg_r(p_j)) \qquad (22)$$

Step 3: Refresh the parameters of the approach, in which lr stands for the learning rate, aims to reduce the impact of each basic model on the result and prevent the framework from being overfit.

$$f_r(p) = f_{r-1}(p-1) + lr * pg_r(p) \quad (23)$$

To combine these two algorithms for energy consumption prediction, we can use a technique called stacking. In stacking, the predictions from multiple machine learning models are combined into a single model that produces more accurate predictions than any of the individual models.

The overall workflow for using Random Forest Regression for energy consumption prediction:

1) Collect and preprocess energy consumption data, including relevant features such as temperature, humidity, and time of day.

2) Create training and validation groupings for the data.

3) Create a Random Forest Regression model using the training set, and then generate forecasts by applying it to the testing set.

4) Create a Gradient Boosting Regression model using the training set, and then generate forecasts by applying it to the testing set.

5) Combine the predictions from the Random Forest and Gradient Boosting models using stacking.

6) Analyze the stacked model's results on the test set.

7) Fine-tune the hyperparameters of the models to improve their performance, if necessary.

8) Once the models are trained and optimized, they can be used to forecast energy usage based on newly acquired data.

In the Random Forest Regression algorithm, numerous decision trees are built, and the predictions from each tree are combined to produce the final forecast. A randomized subgroup of the training information and a randomly selected portion of the features are used for training every decision tree. The algorithm then averages the predictions of all the decision trees to get a final prediction. Mathematically, the forecast for a new input x can be written as Eq. (24).

$$p = \frac{1}{Nd} * \sum_{j=1}^{Nd} p_j(y) \quad (24)$$

where, Nd symbolizes the amount of decision trees along with $p_j(y)$ symbolizes the prediction of the jth decision tree for the input y. The algorithm also constructs multiple decision trees, but instead of averaging their predictions, it combines them using gradient descent. In this algorithm, each decision tree is trained on the residuals (the discrepancy among the prior tree's anticipated result along with the actual output) of the previous tree. Mathematically, the prediction for a new input y can be written as Eq. (25).

$$p = \sum_1^M f_m\,(y) \quad (25)$$

where, M represents the number of decision trees, along with $f_m(y)$ is the prediction of the mth decision tree for the input y. Random Forest Regression utilized to predict the energy consumption for a given set of features (such as distance, obstacles, and congestion) and can be integrated into the proposed routing algorithm to reduce the network's consumption of energy as much as possible.

F. Link Failure Prediction

In the integrated approach using honeybee waggle dance and bat echolocation with the Bellman-Ford-Dijkstra algorithm, link failure prediction can be obtained through various methods. One possible approach is using statistical analysis and machine learning algorithms to analyze network data, including node connectivity, traffic load, signal strength, and other relevant parameters. Applying supervised machine learning - SML strategies like Artificial Neural Networks (ANNs) can help predict link failure based on historical data and other network features. Moreover, integrating honeybee waggle dance and bat echolocation can provide additional inputs for link failure prediction. For instance, the honeybee waggle dance can be used to estimate the distance and direction of the destination node, which can help identify potentially congested areas and bottleneck links along the path. Similarly, bat echolocation can be used to detect obstacles and other obstructions that may affect signal strength and link quality.

G. Artificial Neural Networks (ANNs)

The structure and the human brain's operation motivate a computational model called an ‘Artificial Neural Network’. It is made up of several linked, layered processing nodes or neurons. The layer that produces the result generates the prediction result, and the data that is the input layer gets the input data. The hidden layers perform intermediate processing on the input data to extract relevant features and patterns. The input data is represented as a vector x with n components: x = [x_1, x_2, ..., x_n]. Each component x_i represents a feature of the input data. Each neuron in the data layer gets a portion of the input vector as the input information is fed through the input layer. To create an output, each neuron in the layers that are concealed performs a weighted compilation of its inputs along with applying an activation procedure. Where $h_{i,j}$ represents the jth neuron's outcome in the ith buried layer. The activation function usually incorporates nonlinearity through the network structure through a nonlinearity.

The ultimate forecast is generated by feeding the final result of the last layer that is concealed through the output layer. Y_k stands for the outcome of the k-th neuron in the yield layer. The yield layer may have multiple neurons, each corresponding to a different output class or regression value. The weights and biases of the neurons are learned during the training phase of the network. A cost function that calculates the disparity between the outcomes predicted along with the outcome actually produced is what training aims to minimize. Using an optimization method like gradient descent or Adam, the weights are refreshed. In predicting link failure, ANNs can be used to predict the probability of link failure based on historical data and other features such as traffic load, weather conditions, and network topology. The input vector x symbolizes the features of the link; in addition to the output, y represents the probability of link failure. The biased and weighted elements of the neural network are modified to minimize the discrepancy across its projected likelihood alongside the actual possibility of link failure after it has been trained on the data set consisting of previous link failure occurrences. By the present status of the input characteristics, the ANN can be utilized for instantaneously predicting the likelihood of failure of the link once it has been trained. If the predicted probability exceeds a certain threshold, a link failure is predicted, and the network can take appropriate action to reroute traffic and avoid network disruption.

Steps:

1) Data collection: Collect data related to network topology, traffic, and energy consumption. This data includes features such as link distance, bandwidth, traffic volume, and energy consumption.

2) Feature engineering: Determine pertinent features from the gathered information, including statistical measures such as mean, variance, along with standard deviation.

3) Data preprocessing: Normalize the data to a common scale to remove any bias towards features with larger values. Also, divide the data into training, validation, and test sets.

4) Training ANNs: Use the training data to train ANNs for link failure prediction. ANNs are powerful machine learning algorithms that can learn the underlying patterns as well as make predictions based on new input data. In this case, the ANNs will learn to predict link failures based on the input features.

5) Validation and hyperparameter tuning: Verify the trained ANNs' effectiveness using the validation collection and tune the hyperparameters of the ANNs to improve their performance.

6) Testing: Test the final ANN models on the test set to evaluate their performance.

7) Integration with energy consumption prediction: Finally, integrate the ANN-based link failure prediction model with the previously developed energy consumption prediction model that uses Random Forest Regression and Gradient Boosting Regression with honeybee waggle dance and bat echolocation optimized route selection using the Bellman-Ford-Dijkstra algorithm. This integrated model will now be able to predict both link failures and energy consumption and use this information to optimize the process of routing in the network.

Optimized route selection approach In MANET using " honeybee waggle dance and bat echolocation for route selection, Bellman-Ford-Dijkstra algorithm to find the shortest path in a network, By integrating honeybee waggle dance and bat echolocation into the Bellman-Ford-Dijkstra algorithm, we can create an optimized route selection approach that takes advantage of these natural behaviors to improve the efficiency and accuracy of the routing process, ForestBoost Regression is a novel machine learning algorithms that can be used for energy consumption prediction, Artificial Neural Networks provides more accurate and timely predictions of link failure in MANETs, improving the performance and reliability"

Algorithm for Optimized route selection and Link Failure Prediction

1. Data preprocessing:

Input: Energy consumption data (X), network topology data (G), and link failure data (Y)

Output: Normalized energy consumption data (X_{nor}), normalized network topology data (G_{nor}), and binary link failure data (Y_{bin})

The data preprocessing step involves normalizing the energy consumption data X and the network topology data G, and converting the link failure data Y into binary form. Mathematically, we can represent this step as follows:

$$X_{nor} = (X - \text{mean}(X)) / \text{std}(X) \qquad (26)$$

$$G_{nor} = (G - \text{mean}(G)) / \text{std}(G) \qquad (27)$$

$$Y_{bin} = 1 \text{ if } Y > 0, \text{ else } 0 \qquad (28)$$

where, mean(.) and std(.) represent the mean as well as standard deviation of the data, respectively.

2. Honeybee waggle dance and bat echolocation:

Input: Normalized network topology data (G_{nor})

Output: Estimated distance to destination node (d) and estimated angle to destination node (λ)

The honeybee waggle dance and bat echolocation steps use the normalized network topology data G_{nor} to estimate the distance d and angle theta to the destination node. Mathematically, we can represent this step as follows:

d, λ = honeybee_waggle_dance_and_bat_echolocation(G_{norm}) (29)

where, honeybee_waggle_dance_and_bat_echolocation (.) represents the function that estimates the distance and angle using the honeybee waggle dance and bat echolocation techniques.

3. Bellman-Ford-Dijkstra algorithm with link failure prediction:

Input: Normalized network topology data (G_{nor}), estimated distance to destination node (d), estimated angle to destination node (λ), and binary link failure data (Y_{bin})

Output: Shortest route between source and target nodes (P)

The Bellman-Ford-Dijkstra algorithm with link failure prediction step takes as input the normalized network topology data G_{nor}, the estimated distance towards the target node d, the estimated angle to the destination node theta, along with the binary link failure data Y_{bin}, and outputs the shortest route between source and target nodes. This step consists of two sub-steps: Bellman-Ford algorithm and Dijkstra's algorithm.

3.1 Bellman-Ford algorithm:

Input: Normalized network topology data (G_{nor}), estimated distance to destination node (d), estimated angle to destination node (λ), and binary link failure data (Y_{bin})

Output: Distance estimate (d_s[x]) as well as predecessor node (p_s[x]) for each node in the network

The Bellman-Ford algorithm takes as input the normalized network topology data G_{nor} , the estimated distance to the destination node d_n, the estimated angle to the destination node theta, and the binary link failure data Y_{bin}, and outputs the distance estimate d_n[x] and predecessor node p[x] for each node x in the network. Mathematically, we can represent this step as follows:

```
Initialize $d_n$[x] = infinity, $p_s$[x] = null for all x in G
Set $d_n$[source] = 0
For i = 1 to |X|-1 do
    For each edge (u, x) in G do
        if $Y_{bin}$[u,x] == 0 then
        if $d_n$[x] + $w_t$(u,x) < $d_n$[x] then
        $d_n$[x] = $d_n$[x] + $w_t$(u,x)
        $p_s$[x] = u
```

where, |V| is the number of nodes in the network.

IV. Performance Analysis

An arbitrary collection of origins nodes was replicated with sizes varying from 25 to 125 nodes using the simulation platform, as shown in Table II, to test the efficacy of the suggested mechanisms. These node sources were configured to transmit CBR data packets, which are the nodes' transmission range of 250 seconds, at arbitrary speeds increasing from 25 m/s to a maximum of 30 m/s between arbitrary standstill intervals ranging from 0 to 50 s. The RWP technique is used to generate various nodes. 700s of simulation time is adequate to determine network congestion, latency, and complexity. The optimal path parameter, or "OPI," is generated for each alternate route based on the values of parameters like Movement Indication, Network Access, Path Accessibility, as well as Link Duration. In a MANET, the route with the highest Path selection factor is considered for data transfer across the intermediary nodes towards a target node.

To substantiate the effectiveness of our proposed strategy, we recognize the critical importance of employing diverse and representative datasets that encapsulate various scenarios and environmental conditions inherent in MANET operations. Our dataset selection criteria prioritize factors pivotal to MANET performance, encompassing node mobility, signal strength, and network traffic. For node mobility, we intend to incorporate datasets that emulate a spectrum of mobility patterns and scenarios, including different speeds, pause times, and movement trajectories. The datasets related to signal strength will account for real-world fluctuations, considering interference, obstacles, and varying distances between nodes. Additionally, our approach involves simulating diverse network traffic scenarios, encompassing varying loads, sudden spikes, and fluctuations in traffic. This comprehensive dataset strategy aims to ensure the validity and generalizability of our experimental evaluations, evaluating the adaptability and responsiveness of the proposed MANET optimization strategy across a wide array of realistic conditions. The selected metrics for evaluation include node mobility metrics, signal strength metrics, network traffic metrics, and metrics related to topology changes, providing a holistic assessment of our strategy's performance and its real-world applicability.

TABLE II. SIMULATION PARAMETERS

Parameters	Values
Speed of the node	[25-30]m/s
Number of Nodes	25 - 125
Packet size	512 bytes
Simulation Time	700s
Traffic category	CBR
Protocol used for Routing	RIFA
Mobility Model	Random
Pause Time (s)	10s
Wireless Range of Transmission	250s
Area of Simulation	1200m
Node Assignment	Random

A. Discovery Signal Delivery Probability

To evaluate the performance of the proposed approach for optimized route selection in MANETs, one metric that can be used is the discovery signal delivery probability. This metric measures the ability of the routing protocol to deliver packets to their intended destination. Meeting chance refers to the likelihood that a relay entity will move into conversation with a destination entity. According to the algorithm above, the relay node's choice of the transmitted finding signal is based on the probability's variant value combination. The probability that a signal will be transferred successfully via object Rd to the other device Dd is represented by the Prob(Rd, Dd) symbol. The following procedures are provided for upgrading the detection delivery of messages probability Prob(Rd, Dd): As meeting frequency along with contact time grows, the likelihood of acquiring a signal rises. Therefore, whenever the two devices come together, they will transmit a likelihood table based on the provided mathematical framework, updating the discovery signal by employing Eq. (30) and Eq. (31)

$$Prob(Rd,Dd) = Prob(Rd,Dd)_{Prev} +$$

$$(1 - Prob(Rd,Dd)_{Prev}) * Prob_{orig} * z^{\frac{T_{RdDd}}{T_r + T_d/2}} \tag{30}$$

$$T_{RdDd} = \sum_{q=1}^{l} t_{RdDd}(q) = \sum_{q=1}^{l} (t_{RdDd_{end}}(q) - t_{RdDd_{start}}(q)) \tag{31}$$

The entire contact time across the relay unit Rd along with the destination unit Dd. t is indicated by the symbol T_{RdDd}. The end, followed by the qth connection start times link among relay devices Rd along with recipient device Dd are denoted by $t_{RdDd_{end}}(q)$ in addition to $t_{RdDd_{start}}(q)$, respectively. In addition, $T_{Rd} = \sum_{q=1}^{n} t_{Rdq}$ represents the period of time during which the relay unit Rd is in communication with other network nodes, and $T_{Dd} = \sum_{k=1}^{n} t_{Ddk}$. The overall amount of time during which that device Dd is in communication with any other device on the same network is nt_Ddk. The duration of contact among devices Rd along with Dd is measured as $\frac{T_{RdDd}}{T_{Rd}+T_{Dd}/2}$ in terms of the mean impact rate $(T_r + T_d)/2$.

To evaluate the discovery signal delivery probability, the proposed approach can be compared to other existing routing protocols for MANETs using NS-3 simulation tool. The simulations can be designed to mimic real-world scenarios and measure the percentage of packets that successfully reach their destination. To evaluate the ForestBoost Regression algorithm's performance, the proposed approach can be compared to existing algorithms used for energy prediction in MANETs, such as ANNs or Support Vector Regression (SVR). The performance can be evaluated based on the accurateness of the energy prediction along with the energy consumption optimization achieved by the algorithm. Similarly, the ANN-based approach for predicting link failure in MANETs can be compared to other existing approaches, such as probabilistic models or machine learning-based models. The evaluation can be based on metrics such as prediction accuracy and false positive/negative rates.

B. Packet Delivery Rate

The proportion of the amount of packets sent by upper layers in relation to the amount of packets that arrived at the destination is known as the packet delivery rate. This standard represents the degree of the suggested way from a starting point to a target. With faster data packet delivery, the suggested technique becomes more effective. Let PDR stand for the data packet delivery efficiency, which can be calculated by applying Eq. (32)

$$\text{Packet Delivery Rate} = \frac{N_{pr}}{N_{ps}} * 100\% \qquad (32)$$

where, N_{pr} stands for the quantity of received packets whereas N_{ps} for the quantity of transmitted packets. The packet delivery ratio - PDR fluctuation for the protocols RRP [12], CHNN [14], and MAR [17] is shown in Fig. 3. The packet delivery ratio declines as node speed rises.

The proposed strategy falls from 97% to 90%, the RRP from 92% to 83%, and the MAR from 93% to 88%. The proposed method has a more excellent packet delivery ratio than previous protocols. The most dependable path to the destination is selected by the recommended routing protocol. In comparison to other options, the selected path can have the highest energy level, require the least amount of energy, and cover the greatest distance. By doing this, the likelihood of a node failure is lower, and data loss is reduced.

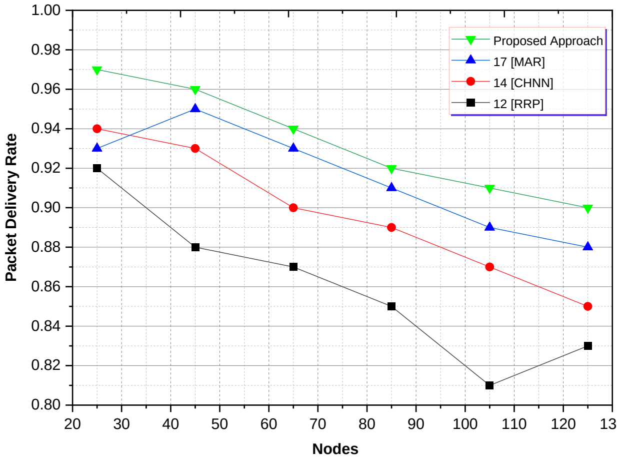


Fig. 3. Variation of packet delivery ratio.

C. Average End-To-End Delay – E2ED

The averaged E2E latency is the duration that it takes for a data packet to arrive which effectively voyage since solitary place to a new. To describe the typical end-to-end delay, we use E2ED. The computation process represented in Eq. (33)

$$E2ED = \frac{1}{Tdp}\sum_{a=1}^{Tdp}\left(T_{pt}(a) - T_{pr}(a)\right) \qquad (33)$$

Fig. 4 compares the suggested approach's average E2E delay to that of the RRP [12], CHNN [14], and MAR [17] range from 10 to 60 minutes. As the amount of time grew, the end-to-end latency shrank. As a result, the suggested method produced a significant latency of roughly 19 ms whenever the pause period was the 60s; however, as the pause time climbed, it dropped due to the lower transportation and likelihood of node failures.

D. Energy Consumption

Energy consumption is the sum of the energy network nodes consumed throughout the scenario. This is accomplished by calculating the energy level of every single node during the end of the trial and accounting for its residual energy. Energy consumption will be represented by the Eq. (34).

$$Econs = \sum_{a=1}^{B}(Ei(an) - Er(an)) \qquad (34)$$

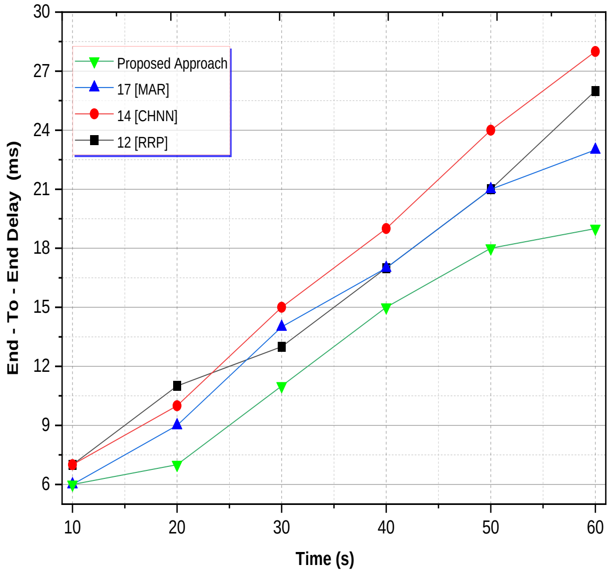


Fig. 4. Average end-to-end delay of the proposed approach.

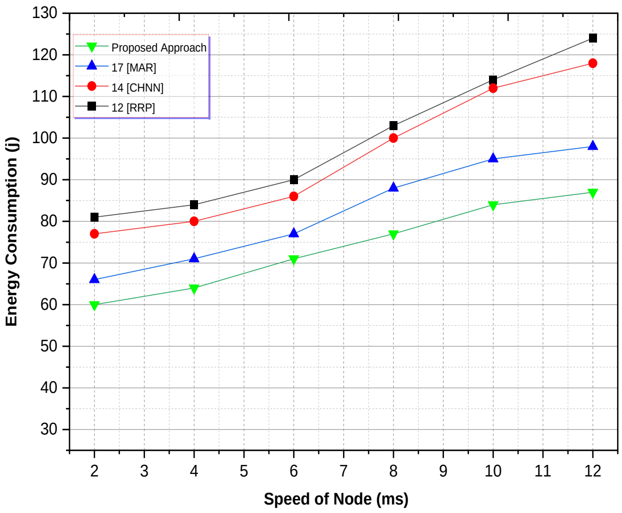


Fig. 5. Energy consumption - node speed.

RRP [12], CHNN [14], and AMR [17] are the three proposed approaches, and the variance in energy usage for each is shown in Fig. 5 (MAR). The energy usage rises as the node speed does. The suggested approach results in an increase of 60 to 87 joules, whereas RRP experiences an increase of 38 to 97 joules, CHNN experiences an increase of 31 to 107 joules, as well as MAR experiences an increase of 35 to 127 joules. The recommended protocol uses less energy than other protocols.

The proposed method groups the routes that eventually reach the desired destination based on their energy levels. The point of origin scatters the data packets while transferring it via pathways with an elevated energy level as well as the normal one in order to equally spread the load on numerous routes. Compared to sending the traffic across just one path, this process uses less energy. RRP, CHNN, and MAR are the recommended approaches, and their energy usage is shown in Fig. 6. In contrast to other protocols, which use more than 30 joules for 10 seconds as well as 90 joules over 60 seconds, the recommended technique uses 29 joules over 10 seconds along with 89 joules over 60 seconds. As a result, the recommended method uses limited energy compared to some alternative methods.

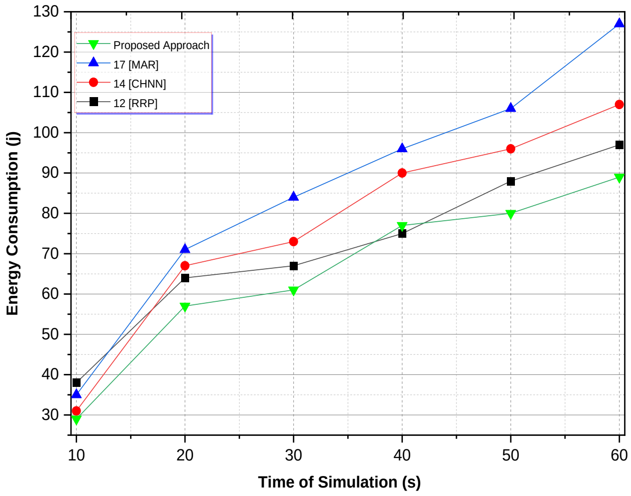


Fig. 6. Energy consumption - simulation time.

The optimization approach proposed for MANET routing offers benefits in improved adaptability, reduced latency, and enhanced energy efficiency. Integrating biomimicry-inspired algorithms and predictive models contributes to dynamic route selection, quicker data transmission, and optimized energy usage. Link failure prediction enhances network reliability. However, challenges include algorithmic complexity, dependence on dataset quality, and considerations for real-world implementation such as hardware constraints and scalability issues. Addressing these limitations is crucial for practical feasibility and efficacy in diverse MANET scenarios.

V. Conclusion and Future work

In conclusion, the optimization of the routing process in MANETs is a crucial task for enhancing the performance and reliability of these networks. The proposed approach, which integrates the biomimicry buzz algorithm with the Bellman-Ford-Dijkstra algorithm, provides a promising solution for this task. By using these behaviors, the nodes can quickly converge on the optimal route to the destination, leading to an efficient and accurate routing solution. Additionally, the use of advanced machine learning techniques such as ForestBoost Regression and ANNs can further optimize the network's energy consumption and predict link failure, improving the network's performance and reliability. These techniques have the potential to revolutionize the routing process in MANETs, enabling them to operate more efficiently and reliably, even in dynamic and unpredictable environments. The proposed approach has the potential to improve the performance and reliability of MANETs, which serves numerous distinct applications in areas such as disaster relief, military operations, and sensor networks. The use of natural behaviors and advanced machine learning techniques is an innovative approach to optimizing the routing process, and further research in this area could lead to even more sophisticated solutions for MANETs.

The proposed optimized route selection approach, coupled with innovative applications of machine learning in predicting energy consumption and link failure, opens avenues for future research and development in the realm of Mobile Ad Hoc Networks (MANETs). The following areas represent potential directions for future exploration:

Further exploration and refinement of biomimicry algorithms, beyond the proposed biomimicry buzz algorithm, could yield enhanced routing strategies. Investigating the application of other bio-inspired algorithms may lead to more efficient and adaptive routing solutions in dynamic MANET environments. Continuous advancements in machine learning techniques can be leveraged to optimize the prediction accuracy of energy consumption and link failure. Exploring deep learning architectures or ensemble methods may enhance the precision of predictions, contributing to more reliable network management. Research into mechanisms for dynamically adapting the proposed algorithm to varying network conditions and node behaviors is essential. The ability to self-adjust based on real-time factors, such as traffic patterns or node density, would enhance the algorithm's adaptability and overall effectiveness.

Incorporating robust security measures within the routing protocol is crucial in MANETs. Future work could focus on integrating advanced security mechanisms to fortify the proposed algorithm against potential security threats, ensuring secure and reliable communication. Translating the proposed strategy from simulations to real-world implementations is a significant future avenue. Evaluating its performance in actual MANET deployments, considering factors like hardware constraints and varying environmental conditions, will validate the practicality and effectiveness of the proposed approach. Enhancing energy prediction models by considering dynamic factors such as node mobility patterns, terrain variations, and changing environmental conditions could further refine energy consumption predictions. Future work might involve developing energy models that adapt to real-time context changes. Extending the experimental evaluation to larger network sizes and assessing the scalability of the proposed approach will provide insights into its performance under more extensive MANET scenarios. Investigating how the algorithm scales with an increasing number of nodes is vital for its practical applicability. Exploring cross-layer optimizations that integrate routing, energy management, and link failure prediction could yield holistic solutions. Collaborative decision-making across multiple protocol layers may enhance the overall efficiency and reliability of MANETs.

References

[1] G. Liu, Z. Yan, and W. Pedrycz, Data collection for attack detection and security measurement in mobile ad hoc networks: A survey, J. Netw. Comput. Appl., 105 (2018) 105–122.

[2] D.-G. Zhang, J.-N. Qiu, T. Zhang, and H. Wu, 'New energy-efficient hierarchical clustering approach based on neighbor rotation for edge computing of IOT, in Proc. ICCCN, Valencia, Spain, 8(1) (2019) 291–295.

[3] S. Glass, I. Mahgoub, and M. Rathod, Leveraging MANET-based cooperative cache discovery techniques in VANETs: A survey and analysis, IEEE Commun. Surveys Tuts., 19(4) (2017) 2640–2661.

[4] S.M. Udhaya Sankar, D. Dhinakaran, C. CathrinDeboral, M. Ramakrishnan, "Safe Routing Approach by Identifying and Subsequently Eliminating the Attacks in MANET," International Journal of Engineering Trends and Technology, vol. 70, no. 11, pp. 219-231, 2022. https://doi.org/10.14445/22315381/IJETT-V70I11P224.

[5] Kacem, B. Sait, S. Mekhilef and N. Sabeur, A New Routing Approach for Mobile Ad Hoc Systems Based on Fuzzy Petri Nets and Ant System, in IEEE Access, 6 (2018) 65705-65720, doi: 10.1109/ACCESS.2018.2878145.

[6] Dhinakaran D, Joe Prathap P. M, "Protection of data privacy from vulnerability using two-fish technique with Apriori algorithm in data mining," The Journal of Supercomputing, 78(16), 17559–17593 (2022). https://doi.org/10.1007/s11227-022-04517-0.

[7] Udhaya Sankar, S.M., Christo, M.S., Uma Priyadarsini, P.S. Secure and Energy Concise Route Revamp Technique in Wireless Sensor Networks, Intelligent Automation and Soft Computing, 35(2) (2023) 2337–2351.

[8] R. Rajeswari, A Mobile Ad Hoc Network Routing Protocols: A Comparative Study, in Recent Trends in Communication Networks. London, United Kingdom: IntechOpen, (2020). doi: 10.5772/intechopen.92550.

[9] Sofian Hamad and Taoufik Yeferny, Routing Approach for P2P Systems Over MANET Network, IJCSNS International Journal of Computer Science and Network Security, vol. 20, no. 3, 2020.

[10] D. Dhinakaran and P. M. Joe Prathap, "Preserving data confidentiality in association rule mining using data share allocator algorithm," Intelligent Automation & Soft Computing, vol. 33, no.3, pp. 1877–1892, 2022. DOI:10.32604/iasc.2022.024509.

[11] V. Ramesh, P. Subbaiah, and K. S. Supriya, Modified DSR (Preemptive) to reduce link breakage and routing overhead for MANET using Proactive Route Maintenance (PRM), Glob. J. Comput. Sci. Technol., 9(5) (2010) 124–129.

[12] E. O. Ochola, L. F. Mejaele, M. M. Eloff and J. A. van der Poll, Manet Reactive Routing Protocols Node Mobility Variation Effect in Analysing the Impact of Black Hole Attack, in SAIEE Africa Research Journal, 108 (2) (2017) 80-92, doi: 10.23919/SAIEE.2017.8531629.

[13] T. D. Nguyen, J. Y. Khan, and D. T. Ngo, A distributed energy-harvestingaware routing algorithm for heterogeneous IoT networks, IEEE Trans. Green Commun. Netw., 2 (4) (2018) 1115–1127.

[14] H. Yang, Z. Li, and Z. Liu, A method of routing optimization using CHNN in MANET, J. Ambient Intell. Humanized Comput., 10(5) (2019)1759–1768.

[15] Dhinakaran, D., Selvaraj, D., Udhaya Sankar, S.M., Pavithra, S., Boomika, R. (2023). Assistive System for the Blind with Voice Output Based on Optical Character Recognition. Lecture Notes in Networks and Systems, vol 492. Springer, Singapore. https://doi.org/10.1007/978-981-19-3679-1_1.

[16] Montoya, C. GuØret, J. E. Mendoza, and J. G. Villegas, ''A multi-space sampling heuristic for the green vehicle routing problem,'' Transp. Res. C, Emerg. Technol., 70 (2016) 113–128.

[17] Chalew Zeynu Sirmollo, Mekuanint Agegnehu Bitew, Mobility-Aware Routing Algorithm for Mobile Ad Hoc Networks, Wireless Communications and Mobile Computing, (2021).

[18] V. V. Mandhare, V. R. Thool, and R. R. Manthalkar, QoS routing enhancement using metaheuristic approach in mobile ad-hoc network, Comput. Netw., 110 (2016) 180–191.

[19] J. Yang, M. Ding, G. Mao, Z. Lin, D.-G. Zhang, and T. H. Luan, Optimal base station antenna downtilt in downlink cellular networks, IEEE Trans. Wireless Commun., 18(3) (2019) 1779–1791.

[20] Taha, R. Alsaqour, M. Uddin, M. Abdelhaq, and T. Saba, Energy efficient multipath routing protocol for mobile ad-hoc network using the fitness function, IEEE Access, 5 (2017) 10369–10381.

[21] H.-C. Liu, J.-X. You, Z. Li, and G. Tian, Fuzzy Petri nets for knowledge representation and reasoning: A literature review, Eng. Appl. Artif. Intell., 60 (2017) 45–56.

[22] S. K. Das and S. Tripathi, Intelligent energy-aware efficient routing for MANET, Wireless Netw., 24(4) (2018) 1139–1159.

[23] S. K. Das and S. Tripathi, Energy efficient routing formation algorithm for hybrid ad-hoc network: A geometric programming approach, Peer-to-Peer Netw. Appl., pp. 1–27, 2018.

[24] Kai, Cui. 'An Ad-hoc Network Routing Algorithm Based on Improved Neural Network Under the Influence of COVID-19'. Journal of Intelligent & Fuzzy Systems, vol. 39, no. 6, pp. 8767-8774, 2020.

[25] N. Shah, H. El-Ocla and P. Shah, "Adaptive Routing Protocol in Mobile Ad-Hoc Networks Using Genetic Algorithm," in IEEE Access, vol. 10, pp. 132949-132964, 2022.

[26] S. Sridhar, V. Nagaraju, B. R. Tapas Bapu, R. Shankar and R. Anitha, "Trusted and Optimized Routing in Mobile Ad-Hoc Networks Emphasizing Quality of Service," Applied Mathematics & Information Sciences, vol. 12, no. 3, pp. 655-663 (2018).

[27] R. Alsaqour, S. Kamal, M. Abdelhaq and Y. Al Jeroudi, "Genetic algorithm routing protocol for mobile ad hoc network," Computers, Materials & Continua, vol. 68, no.1, pp. 941–960, 2021.

[28] Jeena Jacob I.Philip Darney, "Artificial Bee Colony Optimization Algorithm for Enhancing Routing in Wireless Networks," Journal of Artificial Intelligence and Capsule Networks, vol. 3(1), pp. 62-71, 2021.

[29] P. E. Irin Dorathy and M. Chandrasekaran, "Ant-based energy efficient routing algorithm for mobile ad hoc networks," Intelligent Automation & Soft Computing, vol. 33, no.3, pp. 1423–1438, 2022.

[30] Husnain, G.; Anwar, S.; Sikander, G.; Ali, A.; Lim, S. A Bio-Inspired Cluster Optimization Schema for Efficient Routing in Vehicular Ad Hoc Networks (VANETs). Energies 2023, 16, 1456.

[31] S. Sarhan and S. Sarhan, "Elephant Herding Optimization Ad Hoc On-Demand Multipath Distance Vector Routing Protocol for MANET," in IEEE Access, vol. 9, pp. 39489-39499, 2021.

[32] Muhammad Fahad, Farhan Aadil, Zahoor-ur- Rehman, Salabat Khan, Peer Azmat Shah, Khan Muhammad, Jaime Lloret, Haoxiang Wang, Jong Weon Lee, Irfan Mehmood, Grey wolf optimization based clustering algorithm for vehicular ad-hoc networks, Computers & Electrical Engineering, Vol. 70, pp. 853-870, 2018.

[33] Sengathir Janakiraman, A Hybrid Ant Colony and Artificial Bee Colony Optimization Algorithm-based Cluster Head Selection for IoT, Procedia Computer Science, vol. 143, pp. 360-366, 2018.

[34] Charan Pote, Srushti Gulhane, Poorva Rahangdale, "Optimization of Routing Protocol for MANET using BAT Optimization Algorithm," International Journal of Creative Research Thoughts, Vol. 9, no. 5, pp. 212-221, 2021.

[35] A. Junnarkar and A. B. Bagwan, "Novel Quality of Service (QOS) Improvement Routing Protocol for MANET Using Ant Colony Optimization," 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, pp. 1-6, 2017.

Conceptualizing an Inductive Learning Situation in Online Learning Enabled by Software Engineering

Ouariach Soufiane[1], Khaldi Maha[2], Khaldi Mohamed[3]
A Research Team in Computer Science and University Educational Engineering-
Higher Normal School of Tetouan, Abdelmalek Essaadi University, Morocco[1, 3]
Rabat Business School-Rabat International University, Rabat, Morocco[2]

Abstract—**Our work highlights the importance of adopting a systematic and methodical software engineering approach to the development of information technology projects for e-learning. We place particular emphasis on conceptualizing pedagogical scenarios and an inductive online learning situation. To ensure effective management of the information systems development process, we applied instructional design principles and adopted the 2TUP process, a refined version of the Rational Unified Process (RUP) suitable for projects of all sizes. To provide a visual representation of the system architecture and inform instructional design decisions, we used the Unified Modeling Language (UML) to create class, use case, activity, and sequence diagrams. We aim to demonstrate the potential of a structured software engineering approach to creating effective and efficient e-learning systems by conceptualizing an inductive online learning situation and five concrete examples illustrating the system's functionality. Our work underlines the importance of using standardized modeling languages such as UML to facilitate communication between stakeholders and collaboration between instructional designers and software developers.**

Keywords—Software engineering approach; instructional design; conceptualization scenario; online learning situation; inductive approach

I. Introduction

In an IT project, it is imperative to use a structured approach that describes how the project will be carried out. The choice of project management is a decisive phase in the successful completion of the project [1]. A working methodology and development process must be defined, and a project schedule drawn up. In this work, based on the results obtained concerning the conceptualization of our learning activity in a pedagogical scenario according to the modular system, and after having defined our choices in terms of development process and modeling language, we propose a modeling with UML of a system for the conceptualization of a learning situation of a model in online teaching for an inductive approach. Finally, we provide examples of models for our system.

Nevertheless, designing scripting tools for learning activities adapted to blended learning is a complex process that requires a coherent, well-thought-out plan to ensure successful learning [2]. This requires consideration of the pedagogical objectives of the activity, the variety of tools and strategies available, and the needs and interests of the learners [3, 4]. It is important to take this aspect into account, especially when considering a diverse group of distance learners, to be able to offer them appropriate motivational strategies in e-learning systems [5]. Scripting tools are tools used to define and create a series of pedagogical activities in an online environment. They comprise activity diagrams and specification tables to help teachers define pedagogical objectives and organize pedagogical activities in a rational way [6, 7]. On the other hand, teachers can use activity diagrams to clarify pedagogical objectives and organize pedagogical activities, which is an excellent way of planning pedagogical activities and defining tasks in diagrammatic form. A datasheet is used to identify the requirements and methods of a task to be performed, and to provide detailed information.

By using both tools, online instructional storyboard designers can approach their scripts in an organized and rational way. These tools can provide an organized approach to the scriptwriting process, which many people find useful when learning. Within the intricate cycle of a pedagogical scenario in a learning situation lies the revelation of "The four Types of Scripts in an Online Learning Situation" [8]. This paints the picture with four unique stages and related activities. Initiation is up first. Presentation lies in presenting content, performing diagnostic evaluations, and delivering remedial backstage support. The next stop on this cycle is Conceptualization wherein learners dive into tasks such as expounding concepts, using deductive or inductive approach, and arranging information, Group discussions, the practical application of knowledge, and continuous assessment - these activities make up the meat of the Objectivation phase. In the last phase, the Transfer Phase, a pivotal stage in the pedagogical framework is reached. During this phase, a detailed examination of case studies takes place, allowing for a comprehensive analysis. Thorough summative assessments serve as a fundamental cornerstone in evaluating the overall learning outcomes. This phase, aptly named the Transfer Phase, sets the stage for the application of acquired knowledge and the cultivation of practical competencies.

Following having observed the stages of the learning cycle outlined hitherto, conceptualization comprises the secondary phase. It entails structuring the progression from action to implementation, founded upon hindrances encountered throughout situational undertakings. In the end, this stage fosters meaning-making surrounding specialized knowledge, its implementation, and integration in skill development. The stage encompasses two scenarios contingent on the adopted approach. Indeed, contingent on context and situation, proposed learning activities necessitate opting between deductive and inductive approaches. Based on this digital pedagogical framework, presently we examine the second phase concerning conceptualization, with particular focus on the inductive approach.

As part of our comprehensive study, we delve into the second phase of conceptualization, especially the enigmatic inductive approach. As part of our research, we first examine a comprehensive theoretical framework, wherein we reveal the essence of the inductive approach, followed by the presentation of a meticulously constructed scenario designed for the conceptualization of a learning situation of a module in online teaching for an inductive approach.

As we progress, we engage in a compelling discourse on the development process and modeling language employed in our study. We embrace the 2TUP process as a symbol of structural integrity and efficiency while harnessing the potential of UML (Unified Modeling Language). Regarding the modeling system, we venture into the static view, showcasing the class diagram and use case diagram. Further into the dynamic view, we present the modeling of activity and sequence diagrams.

The subsequent phase pertains to the prototyping stage, wherein we materialize our findings through visually perceptible manifestations. Within this context, we present a series of five figures that serve as graphical representations of our system. The initial illustration portrays the authentication model, followed by compelling overviews of the learner dashboard and the teacher dashboard. Additionally, we unveil mock-ups of the wiki and chat pages, thoughtfully presented in both light and dark modes.

II. Theoretical Framework

A. *Scenario for the Conceptualization of a Module Learning Situation in e-Learning for an Inductive Approach*

The inductive approach necessitates proceedings from the explicit to the generalized study in [9]. It represents a scientific technique for deriving broad conclusions from individual premises. This enables learners to discover conceptual significance throughout learning undertakings experimentally. It permits transitioning from actual or concrete observations and examinations to more generalized perspectives. Generalization (objectivization) encourages learners to depict applied methodologies and designate involved operations, drawing upon metacognition, while prompting them to critically and holistically examine their knowledge structures [10]. This process of interdisciplinarity catalyzes the cultivation and reliance on meta-cognitive skills. It fosters a coherent merging of all targets attained, employing dialogue to compile and summarize outcomes. Through generalization, learners explicate individualized scholarship techniques and procedures in aim-related terminologies that can be comprehended and applied more universally [11]. Fig. 1 exhibits an instance of an activity scheme conceptualized for an inductive pedagogical approach concerning a learning scenario founded on the modular system, composed of three systems: input system, Learning system, and output system.

The proposed model is interpreted according to its three systems as follows:

The input system for inductive conceptualization scenarios presents learning activities through defining objectives, knowledge, and skills learners must master.

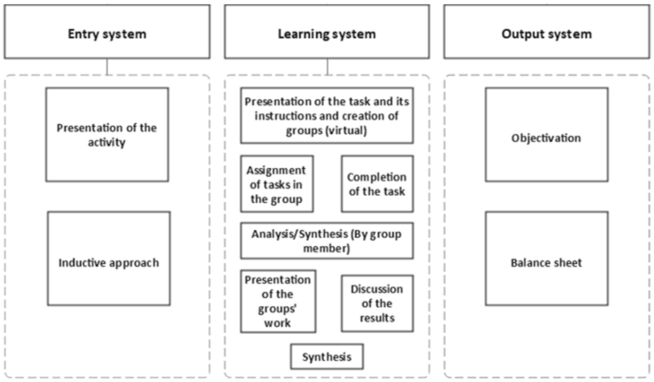

Entry system
Learning system
Output system
Presentation of the activity
Inductive approach
Presentation of the task and its instructions and creation of groups (virtual)
Assignment of tasks in the group
Completion of the task
Analysis/Synthesis (By group member)
Presentation of the groups' work
Discussion of the results
Synthesis
Objectivation
Balance sheet

Fig. 1. Example of a scenario for conceptualizing an inductive approach to a learning situation.

The learning system of the conceptualization activity scenario is based on an inductive approach, involving the transition from specific to general principles [9]. It is consequently prudent to employ the inductive approach, concurring with learner-focused oblique instruction strategies. Indeed, queries, induction, problem resolution, judgment formulation, and discovery are terms interchangeably used to illustrate oblique pedagogy [12, 13]. Indirect teaching advances resourcefulness and interpersonal ability progress. Our system originally proffers a presentation of the undertaking to be fulfilled, contingent on the idea's attributes to be addressed (experiment, problem to be solved, investigation, etc.).

The subsequent step includes enacting the proposed undertaking relative to learner parties' characterizations (autonomous or collaborative exertion), transiting through diverse steps: undertaking accomplishment by organizing and allotting duties in collaborative circumstances, without neglecting time direction; offering explicit experimental constituents for examination by stimulating interrogatives, trials, manipulations and hypothesis development, as well as supplying methods and materials; and ultimately provide feedback by endorsing triumphs, enhancements and self-reformation mediums along with assisting error-discrimination and application. It must be noted undertaking finalization demands supplemental digital asset provision to support learners' labor, and communicative technologies to facilitate interplay between various instructors and learners.

The final step of this educational process involves the examination of the outcomes produced by learners or groups of learners utilizing different communication technologies. Initially, presenting and sharing the results of each individual or group with all participants (teachers and learners) is essential. Following this, the results must be interpreted to validate the empirical findings and compare them with theoretical principles and regulations.

The output system of the scenario of the conceptualization activity of an inductive approach concerns an evaluation of said conceptualization activity as proposed by generalization through engaging metacognition via the manifestation of a model or principles utilizing mathematical formalism.

B. Development Process and Modeling Language

The development process is a decisive factor in the success of a project [14], which outlines the project's phases and determines its fundamental characteristics. Therefore, selecting a development method that is suitable for the project's specific needs and requirements is essential to create a high-quality project that satisfies users' expectations. There are various unified processes, including Extreme Programming (XP) and Rational Unified Process (RUP) [15]. In this particular project, the chosen method is the 2TUP process, which is a hybrid development model that combines the strengths of both RUP and XP and integrates their respective approaches. Our project required a unique approach, and we ultimately decided to utilize the 2TUP process. This process combines the best aspects of both RUP and XP, resulting in a hybrid software development model that integrates their respective approaches. To aid in our project, we employed the unified two-track process, also known as the Y development model. This model follows a split approach, where requirements are studied in one track and technical aspects in the other, before merging them in the lower branch [16]. The Y development model is versatile and suitable for projects of all sizes; it plays a crucial role in managing technical risks and project domains. Additionally, the model addresses the challenges of continuous evolution in information systems by breaking down system analysis along functional and technical axes [17].

Modeling is the design of models. Contingent upon intention and mediums applied. Within a computational discipline, the specification of a phase in constructed informational systems is denoted through data modeling [18]. UML comprises validated engineering optimum processes for modeling sizable, sophisticated schemes [19]. The 2TUP operation relies on UML (Unified Modeling Language) throughout development cycles [20], since its diverse schemas facilitate then simplify the correct modeling of the system at all phases. Verily, UML is depicted as a graphical and textual modeling communicative contrived to apprehend and portray requisites, stipulate, blueprint remedies, and transmit notions [21].

UML unifies object-oriented notation and concepts. It is not alone a representation, but the ideas passed on by the diagrams hold accurate semantics which means in an equivalent approach as text in a language, therefore UML is sometimes depicted as a technique when it truly isn't. UML can be utilized to produce a variety of representations to aid hardware design and development. UML diagrams in use include use case diagrams, class diagrams, sequence diagrams, and activity diagrams [22]. It has two factors: static modeling, which addresses method structure, and dynamic modeling, which concerns the system behavior. Use case diagrams and class diagrams are used to build the basis for static modeling. Dynamic modeling, contrarily, is based on activity and sequence diagrams [23].

UML also unifies the notation required for the various activities in the development process, and hence provides a method for subsequent decision-making, from requirements definition to coding [17]. It is the result of the unification of mature techniques for the analysis and design of large software packages and complex systems.

Consequently, we can conclude that the UML language helps us in all phases of the project, as it offers many advantages in system analysis and design. Consequently, a method is proposed combining UML and Unified Process to guide the realization of object-oriented systems.

III. System Modeling

Given the theoretical data concerning the stages involved in creating an instructional scenario for learning in various disciplines, and more specifically in the case of an online instructional scenario, we analyzed the various components of instructional design for the creation of scenario tools. We based our analysis on the 2TUP development process and the UML modeling language. To create our diagrams, we used StarUML software [24], which also enabled us to make rapid modifications and adjustments to our diagrams.

A. Static View

Static UML (Unified Modeling Language) diagrams are used to model the static structure, i.e., its composition in terms of objects, classes, packages, and so on. There are two diagrams, the class diagram and the use case diagram.

The class diagram is used to represent classes, their attributes, methods, relationships (association, aggregation, composition, inheritance, etc.) a constraint [25].

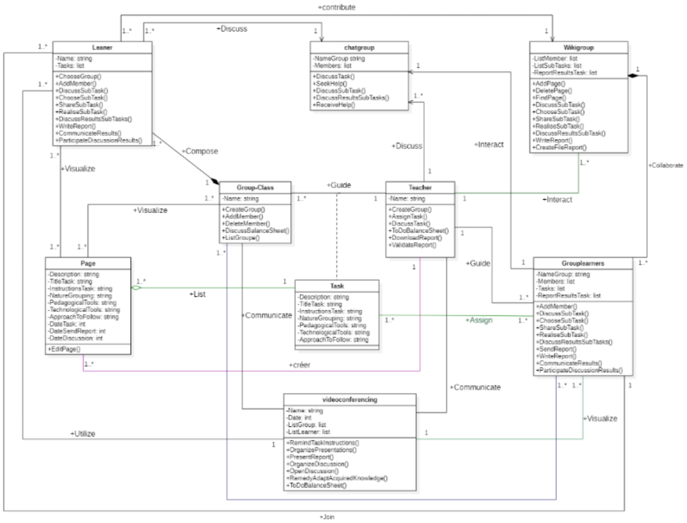


Fig. 2. Class diagram.

- In Fig. 2, which represents the class diagram, we have represented eight classes and one associative class. The eight classes are as follows: Teacher, class-group, learner, group chat, group wiki, learner group, videoconference, and page. The associative class is "task", which results from the relationship between the teacher and the class-group.
- The "Teacher" class enables a teacher to manage groups and assigned tasks. It contains operations such as "create_group()" to create a new workgroup,

"assign_tasks()" to assign tasks to a specific group, "download_reports()" to retrieve learners' reports, and "make_assessment()" to evaluate the group's performance.

- The "Classgroup" class is used to manage workgroups. It contains operations such as "add_member()" to add a member to an existing group, "delete_member()" to remove a member from the group, and "list_group()" to display group members. For example, a group or teacher can use this class to add new members to the group, delete existing members, and display the list of current members.
- The "Task" class is an associative class representing a task assigned to a specific group. It contains attributes such as "description", "title", "TaskInstructions", "GroupNature", "PedagogicalTools", "TechnologicalTools" and "ProcedureToFollow". For example, a task must include a description of the functionality to be implemented, instructions for setting up the functionality, tools, and techniques to be used, and a procedure to be followed to achieve the objective.
- The "Learner" class represents an individual learner. It contains operations such as "choose_group()", "add_member()", "realize_under_task()", and "write_report()", which enable the learner to engage in the learning process. For example, a learner can use this class to choose a specific workgroup, carry out assigned tasks, and write reports on the results obtained.
- The "LearnerGroup" class represents a group of learners. It contains operations similar to those of the "Learner" class, except for an entire group. For example, a group of learners can use this class to discuss assigned tasks, perform tasks together, and communicate results.
- The "Page" class represents a web page containing information about an assigned task. It contains attributes such as "TaskDate", "ReportDate" and "DiscussionDate", along with an "edit_page()" operation for modifying the page. For example, a page can contain information on the functionalities to be implemented, the tools to be used, the deadlines to be met, and the resources available.
- The "Videoconference" class can be used to set up videoconferences to discuss assigned tasks and review progress. It contains operations such as "remind_tache_instructions()", "organize_presentations()", and "make_review()". For example, a group can use this class to organize a videoconference to discuss the tasks assigned, present the results obtained, and evaluate the group's performance by the teacher.
- The "GroupChat" class enables group members to discuss assigned tasks and request assistance. It contains operations such as "discuss_Task()", "request_help()", and "receive_help()". For example, a group can use this class to discuss with the teacher a problem encountered during the completion of a task.
- The "WikiGroupe" class lets you create and manage a wiki for a workgroup. It contains operations such as "add_Page()", "delete_Page()", "find_Page()", "discuss_subtache()", "choose_subtache()", "share_subtache()", "make_subtache()", "discuss_results_subtache()", and "create_report_file()". For example, a group needs to use this class to create a wiki dedicated to their task, add pages for each subtask, discuss subtasks, share useful files and links, and write reports on the results achieved.

Use case diagram represents the interactions between a system and its actors (users, other systems, etc.) [26] in terms of use scenarios and functionalities offered by the system. The use case diagram is often used at the start of the design phase to identify user needs and the main functionalities to be developed. In a use case diagram, actors are represented as external blocks and use cases as ellipses. Relationships between actors and use cases are represented by arrows.

Per the use case diagram shown in Fig. 3, the teacher introduces the activity and corresponding task, forms learner groups, assigns tasks, and oversees assessment and remediation procedures. He assists in group formation and facilitates assessment discussions for the entire class. Within learner groups, participants work collaboratively on the task, delegate sub-tasks, and review and summarize the results. The work is submitted to the educator before the presentation, as per the "include" use case relationship where Case B only transpires upon the execution of Case A.

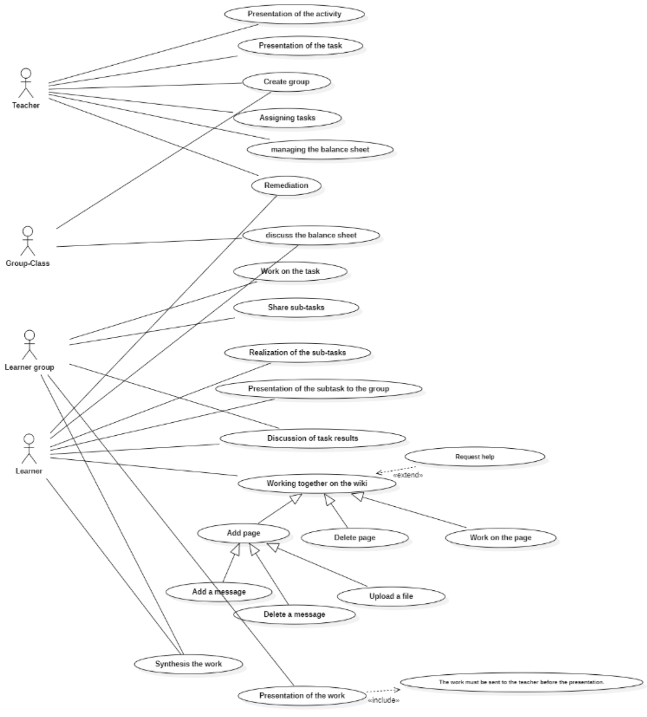


Fig. 3. Use case diagram.

Ultimately, the learner assumes responsibility for conducting sub-tasks within the group, engaging in discussions, collaborating on the wiki platform, and seeking guidance from the teacher if needed, in line with the "extend" use case relationship where Case B can be extended through Case A execution. It is important to note that use case associations merely illustrate interactions between actors and scenarios, not prescribing an accurate sequential order or information flow [27].

B. Dynamic View

UML dynamic diagrams are used to model dynamic behavior, i.e., responses to events and stimuli. In our case, we use the following diagrams sequence and activity diagrams

An activity diagram represents the flow of activities in a process or procedure. It models decision-making, loops, synchronization, and operations specific to task processing.

As shown in Fig. 4, we begin by placing a start node at the center of our activity diagram; this symbolizes the beginning of the activity and is represented by a black circle. This is where the teacher introduces the activity and presents the task and instructions to the group/class. From here, groups are created by the teacher, and the group/class is formed.

Once the task has been assigned to the group, learners can begin to carry it out. If they manage to complete the task successfully, the group of learners starts writing and sending reports. If, on the other hand, the group encounters difficulties, the teacher assists.

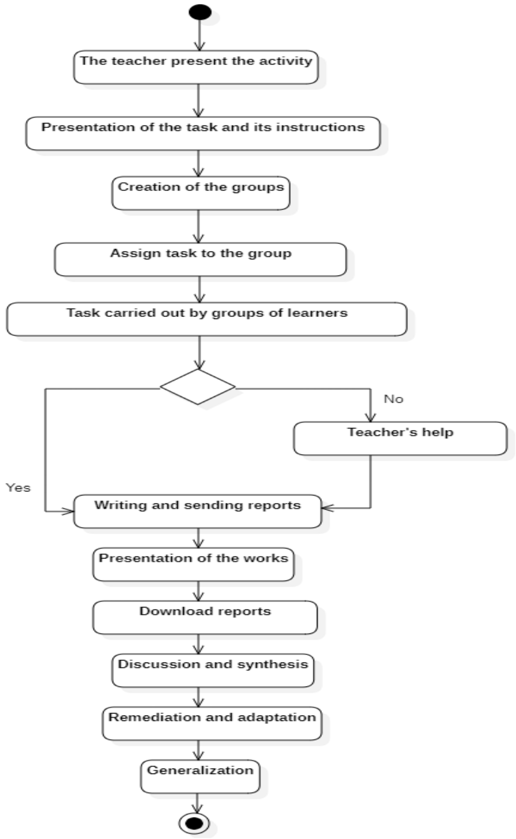

The teacher present the activity
Presentation of the task and its instructions
Creation of the groups
Assign task to the group
Task carried out by groups of learners
No
Teacher's help
Yes
Writing and sending reports
Presentation of the works
Download reports
Discussion and synthesis
Remediation and adaptation
Generalization

Fig. 4. Activity diagram.

After the teacher has uploaded the reports for validation, the groups present their work. Discussion and synthesis then follow, and the teacher can make any necessary adjustments. Finally, the generalization concludes the activity, allowing the teacher to recap the objectives of the activity and encourage learners to continue to work hard and persevere in their learning. The final knot is represented by an outlined black circle.

The sequence diagram shows the interaction between objects over time. It visualizes the sequence of messages exchanged between objects to carry out a given task for a given scenario.

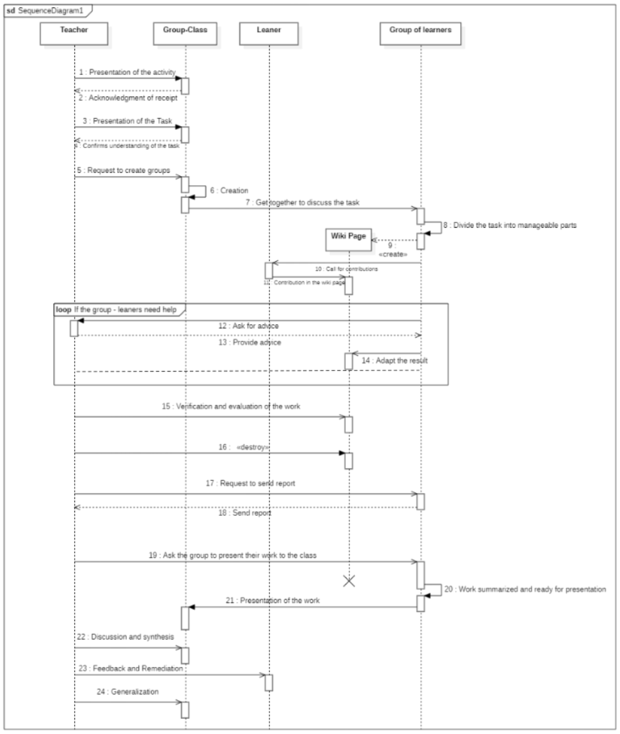

sd SequenceDiagram1
Teacher
Group-Class
Leaner
Group of learners
1 : Presentation of the activity
2 : Acknowledgment of receipt
3 : Presentation of the Task
4 : Confirms understanding of the task
5 : Request to create groups
6 : Creation
7 : Get together to discuss the task
8 : Divide the task into manageable parts
Wiki Page
9 : «create»
10 : Call for contributions
11 : Contribution in the wiki page
loop If the group - leaners need help
12 : Ask for advice
13 : Provide advice
14 : Adapt the result
15 : Verification and evaluation of the work
16 : «destroy»
17 : Request to send report
18 : Send report
19 : Ask the group to present their work to the class
20 : Work summarized and ready for presentation
21 : Presentation of the work
22 : Discussion and synthesis
23 : Feedback and Remediation
24 : Generalization

Fig. 5. Sequence diagram.

The sequence diagram illustrates the interactions between different actors and the main functions of a system as shown in Fig. 5. It shows how the different actors interact to accomplish a task. It starts with the teacher providing knowledge and tasks to the learners and then presenting the task and instructions to the learners. It is inevitable to ask for the creation of groups so that learners can collaborate.

Each member can contribute information to the wiki to contribute to the task.

Our sequence diagram also includes a loop that allows learners to get help from the teacher if they encounter difficulties. By incorporating the teacher's feedback, the group of learners can modify their results to suit their needs. Once the learners have completed their work, the teacher checks and evaluates it. Once the evaluation is complete, the teacher initiates the next phase, requesting reports for the class group. The results of the learners' work can then be discussed and summarized with the teacher. In the event of gaps, the teacher can provide feedback and suggest remedial action if necessary.

In brief, of the UML modeling of our system, once the crucial details had been established, we concentrated on the UML modeling of the system. We created a total of four diagrams, two of them static. The first is the class diagram, made up of eight classes, including an associative class called "task". The second diagram is the use case diagram, which links four actors to fifteen primary tasks and five secondary tasks.

As far as the dynamic diagrams used are concerned, two diagrams were used in this case. The activity diagram was used to visually demonstrate the progression of various actions within a specific process or procedure. In our particular case, we chose to illustrate the progression of the inductive approach in a learning scenario. The second diagram used is the sequence diagram, which delineates the distinct interactions between the various actors involved in the process.

IV. Prototyping

Human-machine interfaces and computer ergonomics are essential elements of modern technology, and can significantly improve user experience and productivity. User-centered design (UCD) is a methodology that prioritizes user needs and preferences in the design process, to create products that are intuitive, efficient, and enjoyable to use [28]. In addition, the use of prototypes will help teams brainstorm and organize their thoughts, leading to more effective problem-solving and collaboration [29]. On the other hand, a functional model or wireframe refers to a diagram used to define the areas and components of a user interface during its design. A wireframe can be created using various methods, such as sketching, paper collage, or digital diagrams.

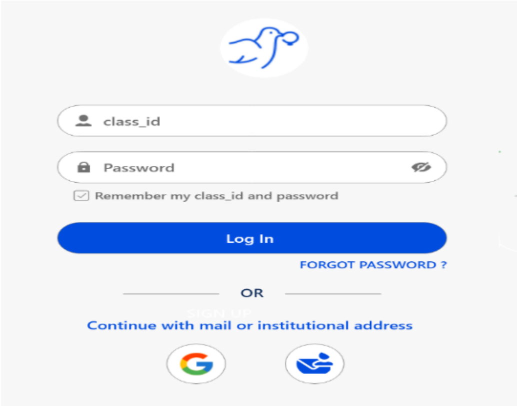
In other words, wireframe modeling is a method used in user experience (UX) design, wireframe modeling enables us to: Identify and address usability issues early in the design process, such as layout. Navigation and content organization. In addition, it allows us to identify potential conflicts between user needs and application capabilities, as well as gaps in the user interface.

The examples in this section illustrate the proposed models. Adobe XD is used to implement these models. Through this work, we propose five models.

A. Authentication Mock-up

The authentication interface enables learners to connect to their workspace using their student ID or institutional address. This provides learners with a simple and secure means of accessing their workspace and collaborating with their peers. Fig. 6 shows a mock-up of the authentication interface enabling users to access the platform.

The username field is designed to accept the learner's ID number or institutional e-mail address, which is linked to their profile. Once the user has entered their credentials, they can click on the validation button to access their workspace. The password field is designed to guarantee the security and confidentiality of the user's account. The password is encrypted and securely stored, and the user is prompted to enter it each time he or she accesses the platform. This is an additional level of security that prevents unauthorized access to the user's workspace. The logo on the authentication interface serves as a visual identifier for the institution and adds a professional touch to the interface. It also reinforces the institution's brand identity and creates a sense of familiarity for users.



class_id

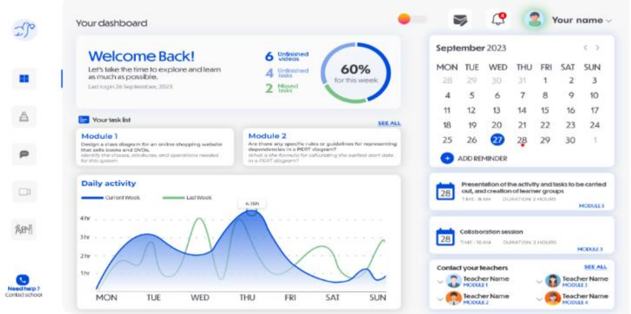
Password

☐ Remember my class_id and password

Fig. 6. Login page.

B. Mock-up for the Dashboard

The dashboard displays the various tasks and activities that have been assigned by teachers, along with the corresponding due dates and deadlines. In addition, learners can see the total time they have spent on the platform, helping them to manage their time effectively. Fig. 7 shows the learner dashboard, which provides an overview of their progress through the course and enables them to monitor their completion status.



Your dashboard

Your name

Welcome Back!

Let's take the time to explore and learn as much as possible.

Last login 26 September, 2023

6 Unfinished videos

4 Unfinished tasks

2 Missed tasks

60% for this week

Your task list

Module 1
Design a class diagram for an online shopping website that sells books and DVDs.
Identify the classes, attributes, and operations needed for this system.

Module 2
Are there any specific rules or guidelines for representing dependencies in a PERT diagram?
What is the formula for calculating the earliest start date in a PERT diagram?

September 2023

MON	TUE	WED	THU	FRI	SAT	SUN
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	1

28 Presentation of the activity and tasks to be carried out, and creation of learner groups
TIME : 8 AM DURATION : 2 HOURS
MODULE 3

28 Collaboration session
TIME : 10 AM DURATION : 2 HOURS
MODULE 3

Contact your teachers

Teacher Name MODULE 1
Teacher Name MODULE 3
Teacher Name MODULE 2
Teacher Name MODULE 4

Fig. 7. Learner dashboard.

The dashboard also includes a calendar that allows learners to see the schedule of upcoming sessions and plan their work accordingly. This feature helps learners stay organized and ensure they don't miss any important sessions or deadlines. In addition, the learner dashboard includes a communication function that allows learners to contact their teachers directly. This function enables learners to ask questions, seek clarification, or provide feedback on course material or teaching. This direct communication channel helps foster a collaborative learning environment and ensures that learners receive the support they need to succeed in the course. In summary, the learner dashboard in Fig. 7 provides an overview of learners' progress in the course, allowing them to track their completion status, view upcoming sessions, communicate with teachers, and access personalized learning resources. The dashboard is designed to be user-friendly and intuitive, making it easy for learners to navigate and use.

C. Mock-up for the Teacher Dashboard

The dashboard is designed to be intuitive and user-friendly, giving teachers easy access to the information they need to teach and manage their courses effectively. Fig. 8 shows the teacher dashboard, which serves as a hub for managing courses, monitoring learners' progress, and communicating with them.

At the head of the dashboard, the teacher can see a planning section for the next session. This section displays the date, time, and duration of the upcoming session, as well as a list of tasks and activities to be covered. The teacher can use this section to prepare for the upcoming session, ensuring that all necessary materials and resources are available. Under the planning section, the teacher can see a folder in which to upload files for each class. This folder enables the teacher to store and organize resources, such as lesson plans, slides, and handouts, in one convenient place. The teacher can also see the progress of the class, including which learners have accessed the material and what tasks have been completed.

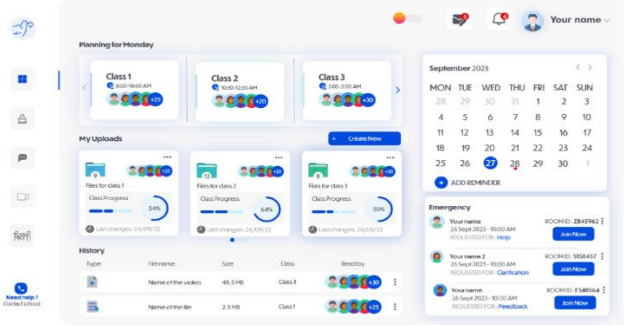


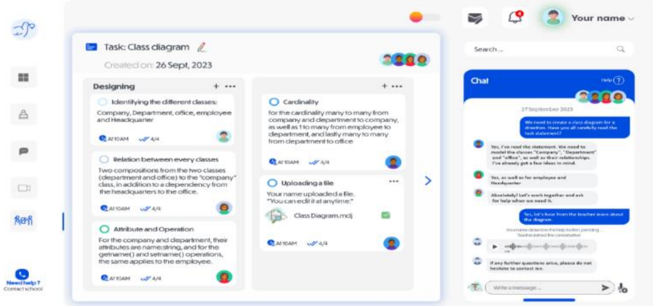
Fig. 8. Teacher dashboard.

In addition, the teacher can consult an emergency section on the dashboard. This section displays a list of learners who need help or feedback on their work. The teacher can quickly identify learners who need help and provide them with the necessary support. At the top of the dashboard, the teacher can also see incoming messages or notifications. This enables them to keep abreast of important announcements or messages from learners, parents, or other teachers. Finally, the teacher can change the look and feel of the dashboard by choosing between a light and dark mode. In this way, teachers can customize their dashboard to suit their personal preferences and working style.

D. Mock-up of Wiki and Chat Pages

The wiki page allows learners to collaborate and share information, files, and resources, enabling them to work effectively together to achieve a common goal. Fig. 9 illustrates a collaborative learning environment in which a group of learners work together to complete a task proposed by a teacher.

Learners use a wiki page to organize their work, adding and modifying the pages needed to complete the task. On the right-hand side of the interface, learners can access a chat function that enables them to communicate with each other in real-time. This feature enables learners to ask questions, request help, and provide feedback to each other, fostering a collaborative and positive learning environment.



Identifying the different classes:
Company, Department, office, employee and Headquarter

Relation between every classes
Two compositions from the two classes (department and office) to the "company" class, in addition to a dependency from the headquarters to the office.

Attribute and Operation
For the company and department, their attributes are name:string, and for the getname() and setname() operations, the same applies to the employee.

Cardinality
for the cardinality many to many from company and department to company, as well as 1 to many from employee to department, and lastly many to many from department to office

Uploading a file
Your name uploaded a file.
"You can edit it at anytime."
Class Diagram.mdj

We need to create a class diagram for a direction. Have you all carefully read the task statement?

Yes, I've read the statement. We need to model the classes "Company", "Department" and "office", as well as their relationships. I've already got a few ideas in mind.

Yes, as well as for employee and Headquarter

Absolutely! Let's work together and ask for help when we need it.

Yes, let's hear from the teacher more about the diagram.

If any further questions arise, please do not hesitate to contact me.

Fig. 9. WikiPage et chat.

The wiki page also allows learners to upload files, such as documents, images, and videos, to support their work. Learners can add new pages to the wiki as required, and they can modify or delete existing pages as necessary. This flexibility enables learners to organize their work in the way that best suits their needs, and promotes effective collaboration.

In one of the many features of the interface, in this case, the chat function, learners can ask the teacher or other experts for help. Learners can use this function to ask questions, request feedback, or seek clarification on any aspect of the task. The teacher or other experts can then respond to learners' requests, providing advice and support where necessary.

E. Mock-up of Dark Mode

Dark mode is often used to reduce eyestrain [30] and preserve battery life on devices with OLED displays [31], as well as to offer users a more cinematic and immersive experience. Dark mode can also be aesthetically pleasing, with many users finding it elegant and modern. In low-light environments, dark mode can also be easier on the eyes, as it reduces the amount of blue light emitted by the screen. Fig. 10 shows an example of dark mode, a display setting that inverts the color palette of a user interface, with a dark background and light text.

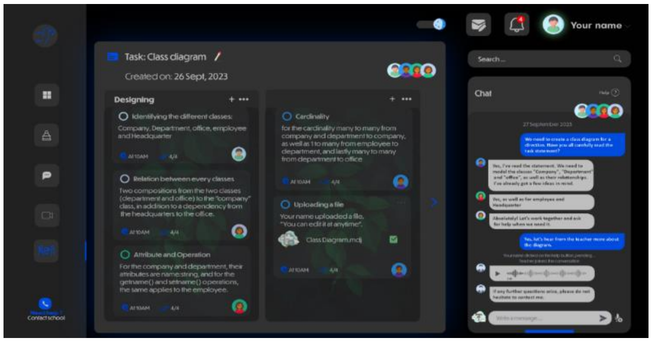


Fig. 10. Dark mode.

To bring this section to a close, we've illustrated five mock-ups. The prototype showcases the transparent authentication process, designed to guarantee secure access to our platform. Users will appreciate the elegant, intuitive interface that simplifies the login procedure.

The second and third prototypes highlight the learner and teacher dashboards, respectively. These well-designed interfaces provide users with an overview of their progress, performance, and personalized learning paths.

Finally, we presented the last two prototypes, which feature a wiki function with chat, a collaborative space for knowledge sharing and discussion. In both prototypes, users can create and edit content, take part in lively discussions, and seek clarification from their peers. Yet the second prototype goes a step further by integrating a visually appealing dark mode option. This feature not only enhances the overall aesthetic experience but also meets the needs of users who prefer a more discreet, immersive interface.

V. Discussion and Limitations

This study proposed a conceptual model for designing inductive learning activities in an online teaching scenario. By grounding the model in established pedagogical frameworks such as the four stages of the learning cycle [8] it aimed to provide a coherent and structured approach.

We applied a structured systems development approach, as previous research has highlighted the importance of structure in projects [32]. As part of our effort to enhance the efficiency and organization of our work, defining a methodology, development process, and schedule was an essential step [33].

We explored conceptualizing an online inductive learning situation using the Unified Modeling Language (UML) for modeling. UML provides a standardized way to visually represent a system's structure and behavior through diagrams like class, use case, activity, and sequence when applied according to established standards [27]. These visual representations, if properly executed facilitate effective communication and collaboration among stakeholders by enabling a shared understanding of system components and functionality [34].

We adopted the 2TUP process to ensure our project followed a systematic, well-defined approach consistent with recommended practices. As described in the relevant literature, 2TUP is rooted in UML and advocates integrating this modeling language throughout the entire development cycle [32]. When applied correctly, UML lends simplicity and clarity to diagrams as an effective means of representation.

By applying this structured systems development approach to analysis and design, we aimed to fulfill system requirements by developing models using object-oriented design and relational database techniques [34]. Creating static and dynamic models allowed us to better understand user requirements and design user-centered solutions.

The prototype we showed demonstrated potential interface screens to refine before full implementation, reflecting best practices. [35]. Learner dashboards and wikis, in particular, depicted collaborative, project-based activities aligned with methods shown by relevant research to motivate e-learners.

In terms of supporting the inductive approach highlighted in existing pedagogical frameworks, the system allows learners to engage with examples and tasks before higher-level concepts are defined. This mirrors the inductive process of drawing inferences from specific cases. Dashboards, wikis, and chat provide a blended environment for facilitating exploration, discussion, and progressive sense-making known from the literature to develop a deeper understanding versus traditional deductive methods [36].

The conceptual models we developed help visualize and plan an online inductive learning scenario according to established pedagogical frameworks. By applying UML and 2TUP, as advocated in relevant sources, we aimed to create a coherent system addressing real user needs, with static and dynamic diagrams clearly illustrating key entities, interactions, and workflows. This type of modeling approach facilitates explaining complex conceptual systems in an organized, visual manner.

The main objective of this article is to discuss the stage of conceptualization, with a specific emphasis on the inductive approach within the pedagogical framework. Although our research delves into the intricacies of this stage, it is important to acknowledge that our focus is limited and does not encompass the entirety of the pedagogical framework. This framework encompasses various other stages, such as the deductive approach, evaluation methods, and other crucial components. While our work is valuable in exploring the inductive approach, it should be viewed as a starting point for further research to incorporate these essential elements. By recognizing these limitations, we can pave the way for future studies that take a more comprehensive and holistic approach to online learning.

VI. Conclusion

In short, our work highlights the fundamental importance of a structured, methodical approach to IT projects, focusing specifically on the design of e-learning pedagogical scenarios for the inductive approach.

After careful evaluation, we decided to adopt the 2TUP (2-Track Unified Process). This approach is well-suited to projects of all sizes and enables us to effectively manage the ongoing evolution of information systems. The division of system analysis into functional and technical aspects offered by 2TUP enabled us to better manage technical risks and ensure effective project management.

We devoted our efforts to UML modeling of our system, creating four diagrams, including two static diagrams. The class diagram presents eight classes, including an associative class called "task". The use case diagram links four actors with fifteen main tasks and five sub-tasks. In contrast, we used two dynamic diagrams: the activity diagram, which illustrates the flow of actions in a process, and the sequence diagram, which describes the interactions between the actors involved.

We aimed to conceptualize an inductive learning situation, with a focus on e-learning. We used UML (Unified Modeling Language) to graphically represent the various components of the system. In turn, we created five concrete mock-up examples to illustrate the system's operation and appearance practically.

In the next stage, our system will undergo a complete development process. It will be developed and rigorously tested to ensure its readiness for deployment. An evaluation will be carried out to measure its effectiveness. Finally, the system will be deployed to its target audience.

REFERENCES

[1] HYVÄRI, Irja. Project management effectiveness in project-oriented business organizations. International journal of project management, 2006, vol. 24, no 3, p. 216-225.

[2] VAUGHAN, Norman D., CLEVELAND-INNES, Martha, et GARRISON, D. Randy. Teaching in blended learning environments: Creating and sustaining communities of inquiry. Athabasca University Press, 2013.

[3] KAHAN, Tali, SOFFER, Tal, et NACHMIAS, Rafi. Types of participant behavior in a massive open online course. International Review of Research in Open and Distributed Learning, 2017, vol. 18, no 6, p. 1-18.

[4] PARK, Ji-Hye et CHOI, Hee Jun. Factors influencing adult learners' decision to drop out or persist in online learning. Journal of Educational Technology & Society, 2009, vol. 12, no 4, p. 207-217.

[5] BOVERMANN, Klaudia et BASTIAENS, Theo J. Towards a motivational design? Connecting gamification user types and online learning activities. Research and Practice in Technology Enhanced Learning, 2020, vol. 15, no 1, p. 1-18.

[6] ANOIR, Lamya, KHALDI, Mohamed, et ERRADI, Mohamed. Personalization in Adaptive E-Learning. In: Designing User Interfaces With a Data Science Approach. IGI Global, 2022. p. 40-67.

[7] KAWTAR, Zargane, MOHAMED, Erradi, et MOHAMED, Khaldi. E-learning adaptive and the tools of screenwriting: a case of collaboration. Global Journal of Engineering and Technology Advances, 2021, vol. 7, no 3, p. 203-212.

[8] MAHA, Khaldi, OMAR, Erradi, MOHAMED, Erradi, et al. Design of educational scenarios of activities in a learning situation for online teaching. GSC Advanced Engineering and Technology, 2021, vol. 1, no 1, p. 049-064.

[9] VARPIO, Lara, PARADIS, Elise, UIJTDEHAAGE, Sebastian, et al. The distinctions between theory, theoretical framework, and conceptual framework. Academic Medicine, 2020, vol. 95, no 7, p. 989-994.

[10] GODDIKSEN, Mads et ANDERSEN, Hanne. Expertise in interdisciplinary science and education. 2014.

[11] MCKEOUGH, Anne, LUPART, Judy Lee, et MARINI, Anthony (ed.). Teaching for transfer: Fostering generalization in learning. Routledge, 2013.

[12] BRUNER, Jerome S. The process of education. Harvard university press, 2009.

[13] JO, Hyun-Jae, CHUNG, Seung-jin, et SATTERFIELD, Debra. Indirect Teaching for All and Autism Spectrum Disorder (ASD) in Design Class (ITAD) Encouraging Their Emotional Empathy. In : Advances in Affective and Pleasurable Design: Proceedings of the AHFE 2016 International Conference on Affective and Pleasurable Design, July 27-31, 2016, Walt Disney World®, Florida, USA. Springer International Publishing, 2017. p. 581-591.

[14] GOVIL, Nikhil et SHARMA, Ashish. Estimation of cost and development effort in Scrum-based software projects considering dimensional success factors. Advances in Engineering Software, 2022, vol. 172, p. 103209.

[15] Sharma, N., & Wadhwa, M. Exsrup: Hybrid software development model integrating extreme programing, scrum & rational unified process. TELKOMNIKA Indonesian Journal of Electrical Engineering, 2015, 16(2), 377-388.

[16] Céret, E., Dupuy-Chessa, S., Calvary, G., Front, A., & Rieu, D. A taxonomy of design methods process models. Information and Software Technology, 2013, 55(5), 795-821.

[17] P. Roques, F. Vallée, "UML 2 in action - From needs analysis to J2EE design," Eyrolles, pp. 382, March 2007. ISBN: 978-2-212-12104-9.

[18] VARENNE Franck, SILBERSTEIN Marc, DUTREUIL Sébastien et al., Modeling & simulating – Volume 2. Epistemologies and practices of modeling and simulation. Éditions Matériologique, "Modelling, simulations, complex systems", 2014, ISBN: 9782919694730. DOI: 10.3917/edmat.varen.2014.01. URL: https://www.cairn.info/modeliser-et-simuler--9782919694730.htm

[19] Weilkiens. Systems engineering with SysML/UML: modeling, analysis, design. Elsevier, 2011.

[20] RAMDA, Amel, HARRAK, Sihem, and BOUAZIZ, Hamida Framer. Development of a Mobile Application for Timetable Management at the University of Jijel. 2019. Doctoral dissertation. Jijel University.

[21] CHARROUX, Benoît, OSMANI, Aomar, and THIERRY-MIEG, Yann. UML 2: modeling practice. Paris: Pearson Education, 2010.

[22] LIU, Xianhong. Identification and check of inconsistencies between UML diagrams. In : 2013 International Conference on Computer Sciences and Applications. IEEE, 2013. p. 487-490.

[23] KHALDI, Maha et ERRADI, Mohamed. Design and Development of an e-Learning Project Management System: Modelling and Prototyping. International Journal of Emerging Technologies in Learning (iJET), 2020, vol. 15, no 19, p. 95-106.

[24] StarUML. (n.d.). StarUML. Retrieved October 14, 2023, from https://staruml.io

[25] ANWAR, Muhammad Waseem, RASHID, Muhammad, AZAM, Farooque, et al. A unified model-based framework for the simplified execution of static and dynamic assertion-based verification. IEEE Access, 2020, vol. 8, p. 104407-104431.

[26] ALERYANI, Arwa Y. Comparative study between data flow diagram and use case diagram. International Journal of Scientific and Research Publications, 2016, vol. 6, no 3, p. 124-126.

[27] ADNAN, Nor Hafizah et RITZHAUPT, Albert D. Software engineering design principles applied to instructional design: What can we learn from our sister discipline?. TechTrends, 2018, vol. 62, p. 77-94.

[28] LOWDERMILK, Travis. User-centered design: a developer's guide to building user-friendly applications. " O'Reilly Media, Inc.", 2013.

[29] DEININGER, Michael, DALY, Shanna R., SIENKO, Kathleen H., et al. Novice designers' use of prototypes in engineering design. Design studies, 2017, vol. 51, p. 25-65.

[30] ERICKSON, Austin, KIM, Kangsoo, BRUDER, Gerd, et al. Effects of dark mode graphics on visual acuity and fatigue with virtual reality head-mounted displays. In : 2020 IEEE Conference on virtual reality and 3D user interfaces (VR). IEEE, 2020. p. 434-442.

[31] EISFELD, Henriette et KRISTALLOVICH, Felix. The rise of dark mode: A qualitative study of an emerging user interface design trend. 2020.

[32] KERZNER, Harold. Project management: a systems approach to planning, scheduling, and controlling. John Wiley & Sons, 2017.

[33] KHALDI, Maha, ANOIR, Lamya, ERRADI, Omar, et al. From Analysis to Development of the E-Learning Project Management System (SGPE). In : Handbook of Research on Scripting, Media Coverage, and Implementation of E-Learning Training in LMS Platforms. IGI Global, 2023. p. 94-132.

[34] CAVIQUE, Luis, CAVIQUE, Mariana, MENDES, Armando, et al. Improving information system design: Using UML and axiomatic design. Computers in Industry, 2022, vol. 135, p. 103569.

[35] Andreasen, M M, C Thorp Hansen, and P Cash. *Conceptual Design: Interpretations, Mindset and Models*. *Conceptual Design: Interpretations, Mindset and Models*. Springer. 2015. https://doi.org/10.1007/978-3-319-19839-2

[36] BENITEZ-CORREA, Carmen, GONZALEZ-TORRES, Paul, et VARGAS-SARITAMA, Alba. A Comparison between Deductive and Inductive Approaches for Teaching EFL Grammar to High School Students. International Journal of Instruction, 2019, vol. 12, no 1, p. 225-236.

The Role of AI in Mitigating Climate Change: Predictive Modelling for Renewable Energy Deployment

Nawaf Alharbe*, Reyadh Alluhaibi
Computer Science Department-College of Computer Science and Engineering,
Taibah University, Madinah, 42353, Saudi Arabia

***Abstract*—This study looks at how AI algorithms like Random Forest, Support Vector Machines (SVM), and Deep Boltzmann Machine (DBM) can be used for predictive modeling to make it easier to use renewable energy sources while reducing the negative effects of climate change. Predictive models based on Artificial Intelligence show possible ways to get the most out of green energy sources, which could lead to fewer carbon emissions. The results of the preliminary studies show that these AI systems can make accurate predictions about how green energy will be made because they are good at making predictions and generalizing. This feature makes it possible to use resources effectively, which improves the reliability of the grid and encourages more people to use green energy sources. Ultimately, employing these AI programs will serve as powerful tools in combating climate change and fostering a more sustainable and eco-friendly environment.**

Keywords*—Renewable energy; climate change; predictive models; and Artificial Intelligence (AI)

I. INTRODUCTION

In the past ten years, Artificial Intelligence (AI) has made huge strides that have changed many different businesses. The fight against climate change is one of the most important and useful ways to use this technology. In this introductory talk, we'll look at how AI algorithms like Random Forest, Support Vector Machine (SVM), and Deep Boltzmann Machine can help lessen the effects of climate change by using predictive modeling to make more people use green energy. Climate change is one of the most important problems of our time.

Most of it is caused by people putting greenhouse gases into the air [1]. The main way that these gases are made is by burning fossil fuels. If we want to stop climate change, we need to quickly make a big switch to green energy. But for renewable energy sources to be used effectively, there needs to be reliable predictive modeling that can figure out the best places, sizes, and types of renewable energy systems to be put. In this situation, AI's amazing ability to predict the future is especially useful.

Artificial intelligence (AI) programs like Random Forest, Support Vector Machine, and Deep Boltzmann Machine have recently become powerful tools for modeling the future. In contrast to standard statistical methods, these algorithms make it possible to understand all of the complex, non-linear interactions in large data sets. They are also good for applications that are based in the real world, like predicting the output of green energy installations as the environment changes [2], because they can handle noise and out-of-the-ordinary data points.

During the training phase, the Random Forest method of ensemble learning is used to make a large number of decision trees. The class that best represents the mean of these individual trees is then output. It has many benefits for predictive modeling in the field of green energy. Two of them are that it can handle high-dimensional spaces and that variables can be related in more than one way. The way of machine learning called "support vector machine" is what SVM stands for. It works by making hyperplanes in a place with a lot of dimensions.

Support Vector Machines (SVM) can be used to make predictions in the field of green energy, especially about things like solar irradiance and wind speed, which are important for making solar and wind power, respectively. The Deep Boltzmann Machine (DBM), which can correctly represent distributions of high-dimensional and complex data, is an example of an artificial neural network that is both random and generative. DBMs are used as a tool in energy dispatch methods because they can predict changes in green energy sources. If these Artificial Intelligence systems were used to model renewable energy sources, it could help fight climate change in a big way.

Using AI could help make the planning and installation of renewable energy systems more efficient and save money. One way to reach this goal is to make predictions about the amount of green energy that can be made more accurate and reliable. The smooth shift to a future powered by clean energy depends on strategic planning, allocating resources, and managing risks. Accurate predictive modeling can help with all of these things. But there are some problems with putting these AI programs into systems for distributing renewable energy.

To fully realize the potential of AI in this area, we need to solve a number of problems [3]. These include the availability and quality of data, the need for computational resources, the need for models to be clear, and the need for AI experts and environmental scientists to work together across different fields. Even though these things have gone wrong, there is reason to be optimistic about the future.

As AI algorithms get better at what they do and become more common, it is likely that they will become much more important in tackling climate change through predictive modeling for the growth of renewable energy sources. As we try to switch to a more sustainable energy system, the combination of Artificial Intelligence (AI) and green energy sources has become a key front in the fight against climate change. So, researchers, policymakers, and people in the field who are working toward a more sustainable future can gain a lot from knowing a lot about these AI tools and how to use them properly.

II. Related Work

There is a review of the literature about how AI can be used in environmental research, especially in the area of renewable energy.

Kaginalkar, Akshara, et al. (2021) [1]: As part of a smart city service, you will figure out how well urban computing can be used to control air quality. This piece is mostly about the Internet of Things (IoT), Artificial Intelligence (AI), and the cloud, as well as how they can be used. The main goals of this study are to keep track of and control the air quality in cities.

Bhaga, Trisha Deevia, et al. (2020) [2]: Using satellite images, make a report about how climate change and drought have affected the surface water sources in sub-Saharan Africa. This study uses remote sensing to look at how climate change affects the water flow in the area.

Floridi, Luciano, et al. (2018) [3]: The goal of proposing the AI4People ethical approach should be to build a society that can take advantage of AI's benefits. In this paper, we talk about the opportunities, risks, guiding principles, and specific ideas for how to move AI forward in a socially responsible way.

Nordgren, Anders (2022) [4]: It looks at the social problems that arise when AI and global warming work together. This analysis looks at how AI can be used in climate science, policymaking, and methods for adapting to climate change.

Freitag, Charlotte, et al. (2021) [5]: Estimates, trends, and rules about the real climate and the revolutionary effects of information and communication technology (ICT) need to be looked at closely. This piece looks at how information and communication technologies (ICT) affect the environment and calls for strict rules and careful assessments.

Allam, Zaheer, and Zaynah A. Dhunny (2019) [6]: There is a link that needs to be looked into between big data, AI, and smart towns. This piece wants to look at how data-driven methods and AI could help build smart cities that are efficient and sustainable.

Schmidt, M. (2020) [7]: It explains the EVOX-CPS idea, which is to retrofit existing buildings with eco-friendly cyber-physical systems to help spread sustainable ways of doing things. In the section of the book about green building design, cloud computing, cyber-physical systems, and the Internet of Things (IoT) are all talked about.

These sources give useful background information on how technology and AI can be used to fight climate change, promote sustainable development, and improve environmental monitoring and control in cities.

TABLE I. Comparative Analysis

Citation	Methodology	Advantage	Disadvantage	Research Gap
Kaginalkar, A. et al.	Review of literature	Integrates urban computing with air quality control.	Limited IoT, AI, and cloud technology integration	Implementing the integrated strategy requires more research.
Bhaga, T.D. et al.	Review of literature	Implementing the integrated strategy requires more research.	Limited to Sub-Saharan Africa.	Remote sensing studies needed in other places.
Floridi, L. et al.	Framework creation	Provides an AI ethical framework	Untested framework.	The framework's practicality and efficacy need further study.
Nordgren, A.	Conceptual analysis	Discusses climate change and AI ethics.	Few empirical data	Ethical challenges and answers need further study.
Freitag, C. et al.	Literature critique	Critically analyzes ICT's climate impact.	Awareness of ICT's climate impact overestimation	Estimates and rules need more research.
Allam, Z. and Dhunny, Z.A.	Conceptual analysis	Examines smart city huge data and AI.	Shows how urban big data and AI integration may benefit	Few case studies
Schmidt, M.	Conceptual analysis	Sustainable growth requires green cyber-physical building systems.	Conceptualizes sustainable building integration	Few empirical studies

III. Methodology

Artificial intelligence (AI) could help fight climate change in a big way by making it easier to use green energy sources in a way that is more efficient and sustainable. One of its main selling points is that it is hard to combine intermittent and fluctuating renewable energy sources, like wind and solar, into the power grid. Predictive modeling is a way to estimate how much power will be made by renewables in the future, considering things like the weather, how people use energy, and the limits of the system.

Predictive modeling is a helpful method. Optimizing the planning and operation of renewable energy systems can help reduce greenhouse gas pollution and increase energy security. This can be done with the help of accurate and reliable predictive models. Yet, it's not easy to work with complex, nonlinear, and high-dimensional data, which makes building models for renewable energy sources a hard task [8]. Also, many different types of renewable energy sources can have a wide range of properties that require a different modeling method for each. So, to get around the problems of modeling green energy and make accurate, generalizable predictions, we need cutting-edge, powerful AI algorithms. These programs must be able to deal with how hard it is to model renewable energy sources. Random forests, support vector machines (SVMs), and deep Boltzmann machines (DBMs) are three of the most popular ways to use AI.

In this study, we put these three AI methods together into a single mixed algorithm [9]. The Random Forest algorithm is a way to learn by putting together the results of several different decision trees. The support vector machine (SVM) is a type of guided learning that tries to find the most statistically significant hyperplane that splits a dataset into more than one class. DBM is an unsupervised learning method [10] that creates a deep generative model of the data using stochastic binary units. The hybrid algorithm takes the best parts of both methods and uses a majority vote to combine the results. We use real-world data sets and the blend algorithm to predict how much wind and solar power will be made in the future.

It has been shown that the hybrid algorithm is better than both the different methods and the other baseline approaches. We show that the hybrid algorithm can accurately predict the variability and uncertainty of green energy sources, and that it is more accurate and reliable than the other methods. Three Artificial Intelligence (AI) algorithms—Random Forest, Support Vector Machines (SVM) [11], and Deep Boltzmann Machine—are used in the plan to fight climate change by using predictive modeling to increase the use of green energy.

A. Random Forest (RF)

The RF algorithm will be used to decide which traits to use and how important they are. This study will help shape plans for putting renewable energy to use in ways that consider climate, regional policies, and available resources, among other things. Climate change can be slowed down by making accurate predictions about how green energy will be used.

The Random Forest (RF) algorithm [12] in AI makes this possible. The RF algorithm uses a lot of data, like weather patterns, how much energy is used, and how well the grid works, to give an accurate assessment of the possibility for renewable energy as explained in Algorithm 1. The deployment plans are based on the results of these analyses.

This foresight promotes the best use of renewable energy sources, which in turn makes us less dependent on fossil fuels and helps slow down climate change. RF is a very flexible machine learning technique [13], so as more data becomes available, it gets better at making predictions. This, in turn, leads to the improvement of methods for deploying green energy [14].

B. Support Vector Machines (SVM)

We will use a support vector machine (SVM) for a regression study to figure out how well green energy systems will work in a wide range of situations. SVM will help make the best predictions of energy yield by modeling complex, nonlinear interactions [15]. So far, so good. AI has been very helpful in the fight against climate change in the area of predictive modeling for the growth of green energy, and SVMs are a key part of this [16].

Algorithm 1:

```
# Set the woodland tree count to 100.

# Define dataset features and labels features = ["solar irradiance", "wind
speed", "temperature", "humidity", "demand"].
labels="solar power output," "wind power output"

# Create training and testing datasets
train_features = split_dataset(features, labels)

# Create an empty tree list forest = []

# For each forest tree in range(n_trees):

  # Randomly sample the training data with replacement sample_features,
sample_labels = bootstrap_sample(train_features, train_labels)

  # Create a decision tree from sample data tree = build_tree(sample_features,
sample_labels)

  # Add tree to woodland.append(tree)

# Define a forest prediction function: predict(forest, test_features):

  # Create an empty list for predictions predictions = []

  # Each test feature in test_features:

    # Create an empty list for tree votes votes = []

    # Forest trees:

      Label = tree.predict(feature)

      # Label votes.append(label)

    # Use majority voting or averaging prediction = aggregate(votes)

    # Add the prediction predictions.append(prediction)

  # Predictions

# Predict test data using the forest predict(forest, test_features)

# Evaluate model performance using some metric score =
evaluate(predictions, test_labels)
```

In the future, when using sustainable energy will be very important, SVM can make predictions about the best places to put new renewable energy infrastructure, like wind farms and solar plants, by looking at weather trends, geographical data, and other factors.

These predictions help place green energy sources in the best places, which makes them work better and make more energy. SVM can also be used to predict how much power can be made from renewables that are already in place [17], considering things like the weather and how often maintenance is scheduled. With the help of this projection, energy suppliers may be able to better meet customer needs if they use less fossil fuels and make the grid more stable.

SVM is a reliable method of machine learning that can help a lot in the fight against global warming. The Algorithm 2 shows how important AI is for helping to build a world that is sustainable.

A Deep Boltzmann Machine (DBM), which can find high-level trends in data, can be used to predict how people will use renewable energy in the future [18]. It will be able to find patterns in the data that has been collected, which will help predict [19] both long-term behavior and possible problems that could stop a lot of people from using green energy. Deep Boltzmann Machines (DBMs), which are a type of generative deep learning model, have been used in many different areas.

Algorithm 2:

```
SVM_Algorithm procedure:

  Initialize Dataset: Load climate change and renewable energy data (wind
patterns, solar irradiance, temperature, humidity, past energy output, etc.).

  Preprocess Data:
    Standardize data.
    Handle missing values.
    Data into Training and Testing sets

  Define SVM Model:
    Select a kernel (linear, polynomial, RBF).
    Set the C parameter (error penalizing/decision boundary margin).
    Choose SVM variant-specific hyperparameters.

  SVM Training:
    SVM-fit the training data.
    Minimize cost function to determine ideal hyperplane
    Store support vectors, hyperplane parameters

  Evaluate the SVM Model: Predict renewable energy generation using the
fitted model on testing data.
    Use metrics like Mean Absolute Error and Mean Squared Error to assess
model performance.

  Good model performance:
    Predict renewable energy generation with SVM.
  Else:
    Change model structure or hyperparameters
    Return to 'Train the SVM Model'.

End Process
```

One of these areas is predictive modeling for the use of green energy sources. DBMs can help AI play its important role in reducing the effects of climate change by giving accurate predictions and insights that can be used to create and implement renewable energy sources [20]. In the next part, we'll talk about how DBMs can be used in predictive modeling to help integrate renewable energy sources as explained in Algorithm 3 [21]. This will help us learn more about how DBMs can be used.

C. Proposed Hybrid Algorithm

Our proposed hybrid method uses the ensemble learning features of Random Forest to successfully use and generalize across several decision trees. Support Vector Machine is used a lot because it can deal with high-dimensional data by finding the best hyperplane [22].

Deep Boltzmann Machine is a deep learning method that gives us another way to describe features. This time, it does this by capturing complex patterns in the data about how to use renewable energy. The goal of combining these three algorithms is to make forecasts more accurate so that they can be used to make decisions about green energy and reducing climate change.

Algorithm 3:

```
# Initialize DBM.
Init DBM (num_layers, num_hidden_units)

# Load renewable energy and climate change dataset
Load dataset

Preprocess the dataset
Dataset preprocessing

# Preprocessed dataset DBM training
Each total_epoch:
  Dataset mini-batches:
    # Sample positive and negative Gibbs. phase
    GibbsSampling(mini_batch, DBM, 'positive')
    negative_phase = GibbsSampling(mini_batch, DBM, 'negative').

    # Update weights and biases via contrastive divergence
    UpdateWeights (DBM)
    UpdateBiases (DBM)

# Predictive modeling after model training
Predicted Renewable Energy Deployment = DBM for each test_data
point.predict(data_point)

# Assess model performance
Performance = EvaluateModel(DBM, test_data).

# Show model performance

# Use the model for renewable energy deployment decisions to reduce climate
change.
MakeDecision(DBM, new_data)
```

1) Hybrid model integration: The best way to combine the results of the Random Forest, Support Vector Machine, and Decision Tree (DBM) models is to use the Decision Tree (DBM) model.

Several ways, like voting, stacking, and weighted average, can be used to combine the results of multiple individual models.

The above combination improves the hybrid algorithm's accuracy, durability, and ability to be used in different situations.

2) Training and optimization: The information is then used to train the hybrid algorithm, and the parameters are optimized in an iterative way.

Cross-validation and grid search are two methods that can be used to find the best hyperparameter values for a given model. As the hybrid algorithm is trained, it learns new knowledge and changes its predictions based on what it has learned.

3) Predictions about how renewable energy will grow: Once the hybrid algorithm has been trained and fine-tuned, it can be used to model how green energy sources will be used in the future.

The program uses climate data, geographic information, and other relevant parameters as inputs to correctly predict how the best renewable energy resources should be used.

The method gives accurate predictions of how renewable energy sources will be used, which is important for stopping climate change as shown below step wise step. To make these predictions, Random Forest, Support Vector Machines, and Decision Trees are all used together.

Step 1: Data preprocessing

- Clean and preprocess input data.
- Training and testing sets.

Step 2: Random Forest Training

- Train a Random Forest model on training data.
- Cross-validation or grid search hyperparameter tuning.

Step 3: SVM Training

- SVM-train the training data.
- Cross-validation or grid search hyperparameter tuning.

Step 4: DBM Training

- Train a Deep Boltzmann Machine on training data.
- Cross-validation or grid search hyperparameter tuning.

Step 5: Ensemble Prediction

- Each test sample:
- Predict renewable energy adoption with Random Forest.
- SVM model renewable energy rollout.
- DBM model renewable energy deployment.
- Weight the forecasts from all three models.

Step 6: Assessment

- Assess the ensemble model's accuracy, precision, recall, and F1-score.

Step 7: Deployment

- Predict real-world renewable energy deployment with the ensemble model.

Step 8: Post-processing

- Analyze the data and offer renewable energy deployment optimization to combat climate change.

IV. Results Analysis

Several AI algorithms can be put together to make a complete and changing predictive model [23] that can be used to improve the way renewable energy is used and how decisions are made. So, they will be very important in reducing the effects of climate change.

First, we'll collect high-quality [24] datasets that include things like weather records, energy usage numbers, greenhouse gas emission totals, and so on, that affect climate change and the spread of renewable energy sources [25, 26]. Next, we'll preprocess these files to remove any information that isn't needed. The data will be checked for any conflicts, outliers, or missing numbers before it is used. Table II and III describe the simulation parameters and datasets.

TABLE II. Simulation Parameter

Parameter	Description	Value Range
Weather Data	Renewable energy environmental factors	Weather history, forecasts
Time Horizon	Simulation length	Months, years
Geographic Scope	Simulated area	Global, country-specific, local
Renewable Energy Sources	Modeled renewable energy sources	Solar, wind, hydro, geothermal, etc.
Energy Consumption	Energy usage forecasts	Megawatt-hours (MWh)
Energy Demand Forecast	Energy demand forecasted by many factors	Megawatt-hours (MWh)
Energy Production Forecast	AI-modeled renewable energy generation	Megawatt-hours (MWh)
Technology Constraints	Renewable energy restrictions	Land availability, transmission lines
Policy and Incentives	Modeled government incentives	Feed-in tariffs, tax credits, subsidies
AI Algorithms	Predictive machine learning algorithms	Neural networks, decision trees, SVMs
Model Validation	Model verification methods	Cross-validation, error analysis
Optimization Objectives	Optimizing renewable energy deployment goals	Maximizing energy generation, cost reduction, emission reduction
Sensitivity Analysis	How input parameters effect results	Weather, policy shifts
Simulation Outputs	Simulation results	Carbon emissions, renewable energy, economic indicators

TABLE III. Datasets Table

Dataset Name	Description
Solar Radiation	Solar radiation history
Wind Speed	Wind speed history
Temperature	Temperature records
Precipitation	Precipitation history
Energy Generation	Renewable energy generation history
Energy Consumption	Energy use history
Economic Indicators	Renewable energy economics
Geographic Data	Location, terrain, and
Environmental Data	Renewable energy environmental factors

Comparison of the performance of the different models: Random Forest, Support Vector Machine, and Deep Boltzmann Machine make up a hybrid algorithm is as shown in Table IV. Simulations and analyses of the table of results for “The Role of AI in Reducing Climate Change: Predictive Modeling for the Deployment of Renewable Energy” is shown in Fig. 1 and Table I also shows comparative analysis.

TABLE IV. PREDICTIVE MODELING FOR THE DEPLOYMENT OF RENEWABLE ENERGY

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	85.2	84.6	86.5	85.5
SVM	81.9	80.3	83.2	81.7
Deep Boltzmann Machine	89.6	90.1	88.2	89.1
Hybrid Algorithm (RF+SVM+DBM)	92.4	92.7	91.8	92.2

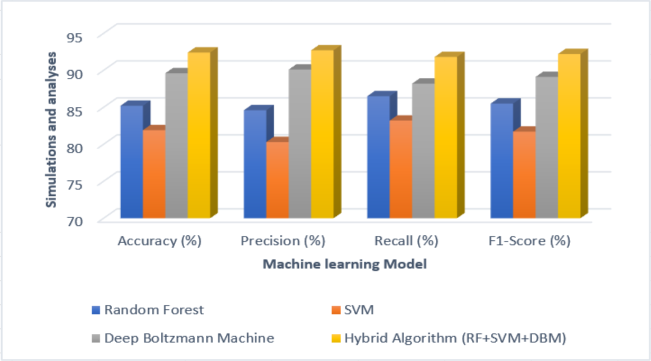


Fig. 1. Comparative analysis of algorithm.

A. Simulation Details

1) Getting information: The dataset used for the simulation included information about past weather trends, energy production, and other important factors for using renewable energy.

2) Design that can be changed: Before the models were trained, a number of methods, such as correlation analysis and information gain, were used to find the most important features for the forecast job.

3) Cross-validation is another method: A five-fold cross-validation approach was used to figure out how well the algorithms worked. After the dataset was randomly split into five parts, different group combinations were used to train and test the model.

4) Measures for judging: Several different factors, such as accuracy, precision, recall, and F1-score, were used to judge how well each algorithm worked.

5) The code for the algorithms was written in Python. The Random Forest, SVM, and Deep Boltzmann Machine were built with the scikit-learn and TensorFlow tools.

6) Making changes to the settings: Grid search and cross-validation were used to fine-tune the hyperparameters and make sure that each method worked at its best.

B. Results Analysis

1) Based on the data that was available, the Random Forest algorithm was able to predict the growth of renewable energy sources with an amazing 85.2% accuracy.

2) The SVM model was as accurate as 81.9% of the time, but it wasn't as good as the Random Forest model.

3) With an accuracy of 89.6%, Deep Boltzmann Machine (DBM) did better than Random Forest and SVM.

4) A hybrid algorithm made up of Random Forest, Support Vector Machines, and Decision Trees had the best accuracy (92.4%), showing how useful it can be to mix different algorithms when using renewable energy sources to make predictions.

5) The hybrid algorithm did better than the individual methods in terms of accuracy, recall, and F1-score, which shows that it can make accurate predictions while striking a good balance between the two.

Overall, the suggested hybrid algorithm was better at predicting the spread of renewable energy than any of the separate algorithms. The Random Forest model, the Support Vector Machine, and the Deep Boltzmann Machine all gave ideas for this method. This shows how AI could help stop climate change by leading decisions about where to put limited energy resources in order to get the most power from renewable sources.

V. DISCUSSION

Climate change is a worldwide problem that needs to be fixed right away. We need to use more alternative energy sources if we want to cut down on greenhouse gas emissions and move toward a more sustainable future. In the last few years, AI has become a powerful tool that can be used to fight climate change. We need to talk about this so we can learn more about how AI, and more especially predictive modeling, can help us use renewable energy sources to fight climate change.

1) Enhancing renewable energy resource assessment: AI-driven predictive modeling could make it much easier and more accurate to evaluate green energy sources. Artificial intelligence systems can look at huge amounts of data to make accurate predictions and maps of where green energy resources are likely to be found.

This knowledge could be about the past climate, the sun's radiation, the speed of the wind, and even the terrain. These forecasting models help lawmakers, financiers, and energy developers find good places to put wind farms, solar farms, and other similar facilities that make energy from renewable sources. By doing this, they can get more done on the projects while lowering the financial risks involved.

2) Optimizing energy generation and storage: Artificial intelligence algorithms could make it easier to manage and run green energy systems by predicting how energy will be used and how much will be produced. By looking at past data on how much power was used, the weather, and how much energy could be stored, machine learning methods can find the best time to generate and store renewable energy to save money and make the grid more reliable.

One way to reach this goal is to make sure that renewable energy is made and kept at the right times. AI can also be used to improve the efficiency of devices that store energy, like batteries. This is done by figuring out the best way to charge and discharge the battery based on how it will be used.

3) Improving energy grid efficiency: AI has a lot of promise to make the energy infrastructure we already have work better and make it easier to add renewable energy sources. Smart grid technologies are based on algorithms that are driven by AI. These algorithms make it possible to accurately predict load, predict congestion, and find the best way to route and distribute energy.

Grid management systems could keep track of and organize different types of renewable energy in real time by using AI. So, less energy will be lost during transfer, and we might even be able to make sure a steady flow of green power.

4) Enhancing energy demand management: Demand response management is the process of changing how much energy is used in reaction to changes in the grid and the availability of renewable energy. Predictive modeling, which is based on AI, can help with this. AI algorithms can predict peaks and valleys in energy use by looking at past data and patterns of consumer activity.

This makes it possible to use automated tools for proactive management of how much energy is used. This method not only lowers the need for fossil fuel power plants during peak hours, but it also lets customers help make the switch to cleaner energy sources by changing how much energy they use based on how much renewable energy is being made.

5) Facilitating policy and investment decisions: When it comes to investing in green energy, AI can help government and business leaders make better decisions. Predictive modeling can help us understand how different green energy sources will work and how much they will cost in the long run.

Policymakers could use these results to make rules and regulations that work better. Using algorithms based on Artificial Intelligence to analyze investment risks, figure out how profitable renewable energy projects might be, and help make financial decisions can help increase investments in the renewable energy industry.

VI. Conclusion

By using AI techniques in predictive modeling for the use of renewable energy, we can learn more about how climate change works, make the best use of energy resources, and move toward sustainable growth. The Random Forest, SVM, and DBM are all examples of these Artificial Intelligence methods.

With these methods, we can use data-driven insights to make good decisions about policy, come up with workable solutions for green energy, and lessen the worst effects of climate change. For successful adoption, however, it's important to keep in mind that AI models aren't a cure-all and must be used with domain knowledge and with the political, social, and economic contexts in mind.

AI methods are very important when it comes to fighting climate change and getting more people to use renewable energy. AI technologies like Random Forest, SVM, and DBM can be used to evaluate, predict, and improve many aspects of how climate and energy combine. Two more types of AI are neural networks and genetic programming. By using AI, we can speed up the switch to renewable energy sources, lower greenhouse gas emissions, and take steps toward a more sustainable and resilient future.

References

[1] Kaginalkar, Akshara, et al. “Review of Urban Computing in Air Quality Management as Smart City Service: An Integrated IoT, AI, and Cloud Technology Perspective.” Urban Climate, vol. 39, Sept. 2021, p. 100972, https://doi.org/10.1016/j.uclim.2021.100972. Accessed 29 Oct. 2021.

[2] Bhaga, Trisha Deevia, et al. “Impacts of Climate Variability and Drought on Surface Water Resources in Sub-Saharan Africa Using Remote Sensing: A Review.” Remote Sensing, vol. 12, no. 24, 21 Dec. 2020, p. 4184, https://doi.org/10.3390/rs12244184.

[3] Floridi, Luciano, et al. “AI4People—an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.” Minds and Machines, vol. 28, no. 4, 26 Nov. 2018, pp. 689–707, link.springer.com/article/10.1007/s11023-018-9482-5, https://doi.org/10.1007/s11023-018-9482-5.

[4] Nordgren, Anders. “Artificial Intelligence and Climate Change: Ethical Issues.” Journal of Information, Communication and Ethics in Society, 10 Feb. 2022, https://doi.org/10.1108/jices-11-2021-0106.

[5] Freitag, Charlotte, et al. “The Real Climate and Transformative Impact of ICT: A Critique of Estimates, Trends, and Regulations.” Patterns, vol. 2, no. 9, Sept. 2021, p. 100340, https://doi.org/10.1016/j.patter.2021.100340.

[6] Allam, Zaheer, and Zaynah A. Dhunny. “On Big Data, Artificial Intelligence and Smart Cities.” Cities, vol. 89, no. 89, June 2019, pp. 80–91, https://doi.org/10.1016/j.cities.2019.01.032.

[7] Schmidt, M. (2020). EVOX-CPS: Turning Buildings into Green Cyber-Physical Systems Contributing to Sustainable Development. In: Ranjan, R., Mitra, K., Prakash Jayaraman, P., Wang, L., Zomaya, A.Y. (eds) Handbook of Integration of Cloud Computing, Cyber Physical Systems and Internet of Things. Scalable Computing and Communications. Springer, Cham. https://doi.org/10.1007/978-3-030-43795-4_13.

[8] Barthelmie, R.J.; Pryor, S.C. Climate Change Mitigation Potential of Wind Energy. Climate 2021, 9, 136. https://doi.org/10.3390/cli9090136

[9] Vijayalakshmi, S., Savita, Durgadevi, P. (2023). AI and IoT in Improving Resilience of Smart Energy Infrastructure. In: Vijayalakshmi, S., ., S., Balusamy, B., Dhanaraj, R.K. (eds) AI-Powered IoT in the Energy Industry. Power Systems. Springer, Cham. https://doi.org/10.1007/978-3-031-15044-9_9

[10] Graditi, G., Buonanno, A., Caliano, M., Di Somma, M., Valenti, M. (2023). Machine Learning Applications for Renewable-Based Energy Systems. In: Manshahia, M.S., Kharchenko, V., Weber, GW., Vasant, P. (eds) Advances in Artificial Intelligence for Renewable Energy Systems and Energy Autonomy. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-031-26496-2_9

[11] Pant, ., Rajawat, A.S., Goyal, S. et al. Machine learning–based approach to predict ice meltdown in glaciers due to climate change and solutions. Environ Sci Pollut Res (2023). https://doi.org/10.1007/s11356-023-28466-0.

[12] Goyal, S.B., Bedi, P., Rajawat, A.S., Shaw, R.N., Ghosh, A. (2022). Smart Luminaires for Commercial Building by Application of Daylight Harvesting Systems. In: Bianchini, M., Piuri, V., Das, S., Shaw, R.N. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes

in Networks and Systems, vol 218. Springer, Singapore. https://doi.org/10.1007/978-981-16-2164-2_24.

[13] Bedi, P., Goyal, S.B., Rajawat, A.S., Shaw, R.N., Ghosh, A. (2022). Application of AI/IoT for Smart Renewable Energy Management in Smart Cities. In: Piuri, V., Shaw, R.N., Ghosh, A., Islam, R. (eds) AI and IoT for Smart City Applications. Studies in Computational Intelligence, vol 1002. Springer, Singapore. https://doi.org/10.1007/978-981-16-7498-3_8.

[14] Rajawat, A.S., Barhanpurkar, K., Shaw, R.N., Ghosh, A. (2021). Risk Detection in Wireless Body Sensor Networks for Health Monitoring Using Hybrid Deep Learning. In: Mekhilef, S., Favorskaya, M., Pandey, R.K., Shaw, R.N. (eds) Innovations in Electrical and Electronic Engineering. Lecture Notes in Electrical Engineering, vol 756. Springer, Singapore. https://doi.org/10.1007/978-981-16-0749-3_54.

[15] Sirmacek, B. et al. (2023). The Potential of Artificial Intelligence for Achieving Healthy and Sustainable Societies. In: Mazzi, F., Floridi, L. (eds) The Ethics of Artificial Intelligence for the Sustainable Development Goals . Philosophical Studies Series, vol 152. Springer, Cham. https://doi.org/10.1007/978-3-031-21147-8_5.

[16] Vijayalakshmi, S., Savita, Genish, T., George, J.P. (2023). The Role of Artificial Intelligence in Renewable Energy. In: Vijayalakshmi, S., ., S., Balusamy, B., Dhanaraj, R.K. (eds) AI-Powered IoT in the Energy Industry. Power Systems. Springer, Cham. https://doi.org/10.1007/978-3-031-15044-9_12

[17] Sharma, N., De, P.K. (2023). Effect of Non-renewable Energy Sources on Climate Change in India—Literature Review and Data Preparation. In: Towards Net-Zero Targets. Advances in Sustainability Science and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-19-5244-9_4.

[18] Y. Shan, J. Hu, Z. Li and J. M. Guerrero, "A Model Predictive Control for Renewable Energy Based AC Microgrids Without Any PID Regulators," in IEEE Transactions on Power Electronics, vol. 33, no. 11, pp. 9122-9126, Nov. 2018, doi: 10.1109/TPEL.2018.2822314.

[19] M. Vašak, A. Banjac, N. Hure, H. Novak, D. Marušić and V. Lešić, "Modular Hierarchical Model Predictive Control for Coordinated and Holistic Energy Management of Buildings," in IEEE Transactions on Energy Conversion, vol. 36, no. 4, pp. 2670-2682, Dec. 2021, doi: 10.1109/TEC.2021.3116153.

[20] M. U. Jan, A. Xin, H. U. Rehman, M. A. Abdelbaky, S. Iqbal and M. Aurangzeb, "Frequency Regulation of an Isolated Microgrid With Electric Vehicles and Energy Storage System Integration Using Adaptive and Model Predictive Controllers," in IEEE Access, vol. 9, pp. 14958-14970, 2021, doi: 10.1109/ACCESS.2021.3052797

[21] N. Sockeel, J. Gafford, B. Papari and M. Mazzola, "Virtual Inertia Emulator-Based Model Predictive Control for Grid Frequency Regulation Considering High Penetration of Inverter-Based Energy Storage System," in IEEE Transactions on Sustainable Energy, vol. 11, no. 4, pp. 2932-2939, Oct. 2020, doi: 10.1109/TSTE.2020.2982348.

[22] G. Mohy-ud-din, K. M. Muttaqi and D. Sutanto, "Adaptive and Predictive Energy Management Strategy for Real-Time Optimal Power Dispatch From VPPs Integrated With Renewable Energy and Energy Storage," in IEEE Transactions on Industry Applications, vol. 57, no. 3, pp. 1958-1972, May-June 2021, doi: 10.1109/TIA.2021.3057356.

[23] Z. Zhao et al., "Distributed Robust Model Predictive Control-Based Energy Management Strategy for Islanded Multi-Microgrids Considering Uncertainty," in IEEE Transactions on Smart Grid, vol. 13, no. 3, pp. 2107-2120, May 2022, doi: 10.1109/TSG.2022.3147370.

[24] H. Novak, V. Lešić and M. Vašak, "Hierarchical Model Predictive Control for Coordinated Electric Railway Traction System Energy Management," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 7, pp. 2715-2727, July 2019, doi: 10.1109/TITS.2018.2882087.

[25] M. E. Zarei, D. Ramirez, M. Prodanovic and G. Venkataramanan, "Multivector Model Predictive Power Control for Grid Connected Converters in Renewable Power Plants," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 10, no. 2, pp. 1466-1478, April 2022, doi: 10.1109/JESTPE.2021.3077953.

[26] Y. Yu, L. Quan, Z. Mi, J. Lu, S. Chang and Y. Yuan, "Improved Model Predictive Control with Prescribed Performance for Aggregated Thermostatically Controlled Loads," in Journal of Modern Power Systems and Clean Energy, vol. 10, no. 2, pp. 430-439, March 2022, doi: 10.35833/MPCE.2020.000834.

An Exploratory Analysis of using Chatbots in Academia

Njood K.Al-harbi, Amal A.Al-shargabi
Department of Information Technology-College of Computer, Qassim University, Qassim, Buraydah, Saudi Arabia

***Abstract*—With the advancement of technology in this era, chatbots have become more than just robots, as they used to conduct time-consuming and labor-intensive routine tasks. Now, it is more than just a robot for routine duties; it interacts and produces like a human. Despite the efficacy and productivity of chatbots like ChatGPT-4 and Bard, there will be significant ethical implications for the academic community, particularly students and researchers. The current study is experimenting with ChatGPT-4 and Bard by producing scientific articles with specific criteria, then applying topic modeling to assess the extent to which the content of the articles is related to the required topic, and verifying references, plagiarism, and the accuracy of the chatbot-generated articles. The results indicated that the content is relevant to the topic, and the accuracy of ChatGPT-4 is greater than Bard. ChatGPT-4 achieved 96%, and the majority of the bibliographies are accurate, whereas Bard achieved 52%, and the majority of bibliographies are incorrect, and some are not available. It is unethical to rely on a chatbot to produce scientific content, despite its accuracy, because it is not as accurate as humans and requires a thorough review of the content it generates. Furthermore, it alters his responses based on the individual he is interrogating, regardless of whether his answers are correct, as he is unable to defend his knowledge.**

***Keywords*—AI; chatbots; ChatGPT; GPT-4; bard; ethics; machine learning; topic modeling**

I. Introduction

Every day, people from all over the world discover and experience new technological miracles due to the rapid development of science in the current era [1]. Who would have predicted that robots would assist humans in completing tasks in a time-saving and efficient manner? ChatGPT (Generative Pre-trained transformer), an artificial intelligence-based chatbot, has recently captivated the attention of many in the tech community. Launched in November 2022, it was developed by OpenAI, a research and publishing company specializing in artificial intelligence (AI). It was built on top of OpenAI's GPT-3.5 and GPT-4 families of large language models (LLMs) and has been fine-tuned using both supervised and reinforcement learning techniques (an approach to transfer learning) [2].

Moreover, Bard is a chatbot. It is based on a large language model (LLM) powered by Google. Improved and lightweight of LaMDA. It is similar to ChatGPT with the difference that it obtains its information from the web directly and it is up to date. It is presently under development and will be enhanced with more capable models over time [3].

Another example of a chatbot operated by Microsoft is the search engine, Bing. It operates as a chatbot assistant that can carry out tasks and can do so through either text or voice conversation using the Open AI concept as the basic and based on ChatGPT and GPT-3.5 [4].

The development of artificial intelligence-based technologies, such as ChatGPT, Bard, and other chatbots, confers both tremendous power and great responsibility. Therefore, ethical concerns must be considered. Although it is beneficial for some routine tasks, such as editing, it has significant bias issues. In addition, her speed in writing research papers poses a threat to scientific integrity.

The current research will examine the impact of chatbots, particularly ChatGPT, on the academic community and how to educate students and researchers on scientific research's ethics and integrity. The following are some of the questions that are currently being researched:

1) *Does* chatting with a chatbot produce useful scientific data for academics? Or does it raise ethical concerns?
2) *What* are the risks behind chatbots?
3) *Is* the chatbot's content real and reliable?

To answer the questions of the research, certain objectives must be met, including:

1) *Comparison* study, of how Google Bard and ChatGPT work and other chatbots.
2) *Experiment* with different chatbots (GPT-4 and Bard).
3) *Check* the bibliography's credibility.
4) *Request* chatbots to generate unique articles and then check for plagiarism.
5) *Discuss* the ethical issues.

The significance of the research is rooted in the keeping of scientific research's integrity and the provision of education to students and researchers regarding the ethics of scientific research. Thus, the responsible and ethical utilization of AI tools is essential within the context of academic research and publication. Furthermore, it is important to consider copyright, authorship, and proper citation of information sources.

The subsequent sections of the paper are organized as follows: Section II provides an overview of the related work, Section III outlines the methodologies and materials employed for exploration, Section IV presents the result of this research, Section V presents a discussion of results, and finally, Section VI presents the conclusion and outlines recommendations for future research.

II. Related Works

This section will present a brief history of chatbots, followed by an explanation of GPT models. Thus, will discuss the ethical issues with it, and conclude with a review of recent research on the topic to examine its impact in academia.

A. Chatbots Background

Alan Turing thought in 1950 whether a computer program could converse with a group of individuals without them recognizing that their interlocutor was artificial. Many consider this question, which is dubbed the Turing test, to be the generative concept of chatbots [5].

In 1966, the first chatbot with the moniker ELIZA was created. Due to its limited knowledge, ELIZA can only discuss a specific domain of topics, which is one of its disadvantages. Additionally, it cannot maintain lengthy dialogues and cannot learn or discover context from them.

In 1988, Jabberwacky was built, making it the first chatbot to employ AI [6, 7]. AI affects our daily lives and activities through many applications and advanced devices, called intelligent agents, which perform a variety of tasks [7]. A chatbot is a program with AI and a paradigm of human-computer interaction (HCI)[8]. It employs natural language processing (NLP) and sentiment analysis to communicate with humans or other chatbots in human language via text or speech [9]. Chatbots are beneficial in many fields, including education, business, e-commerce, health, and entertainment, in addition to entertaining people and simulating human interaction [10].

In November 2022, ChatGPT version 4, a chatbot with extraordinary writing abilities, caused a sensation, particularly in academic circles. Some researchers list him as an author in their academic papers. However, Nature and Science has stated that he cannot be designated as an author under the current legal system because he is not a human being and the text generated by the chatbot cannot be protected by copyright. Even though the most recent version of ChatGPT is advanced, search ethics prohibit its inclusion as an author in a search [7, 11].

B. GPT Models

ChatGPT utilizes a variant of the GPT model which has been trained to respond to questions using a massive dataset [2]. ChatGPT generates responses to text inputs using natural language processing (NLP). GPT models are based on the transformer architecture, which was presented in a paper [12] as a neural network architecture. The architecture of ChatGPT is extremely complex and consists of multiple layers of neurons. The model consists of an encoder and a decoder that work together to generate responses to diverse user inputs. The encoder processes the input text to generate a sequence of hidden states, which are then passed to the decoder. The decoder then employs these hidden states to generate token-by-token output text, a process known as autoregression [13,14]. The processes of ChatGPT are in Fig. 1.

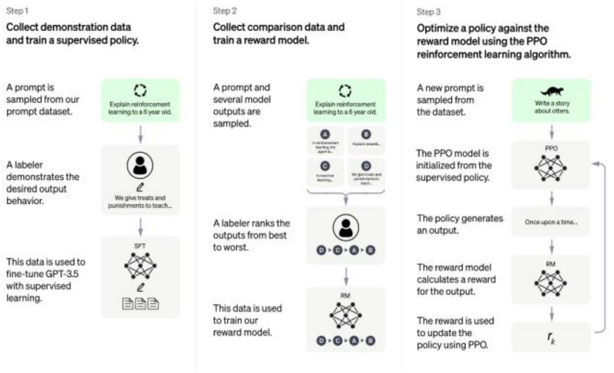

Step 1
Collect demonstration data and train a supervised policy.
A prompt is sampled from our prompt dataset.
Explain reinforcement learning to a 6 year old.
A labeler demonstrates the desired output behavior.
We give treats and punishments to teach...
This data is used to fine-tune GPT-3.5 with supervised learning.
SFT
Step 2
Collect comparison data and train a reward model.
A prompt and several model outputs are sampled.
Explain reinforcement learning to a 6 year old.
A
B
C
D
A labeler ranks the outputs from best to worst.
D > C > A > B
This data is used to train our reward model.
RM
D > C > A > B
Step 3
Optimize a policy against the reward model using the PPO reinforcement learning algorithm.
A new prompt is sampled from the dataset.
Write a story about otters.
The PPO model is initialized from the supervised policy.
PPO
The policy generates an output.
Once upon a time...
The reward model calculates a reward for the output.
RM
The reward is used to update the policy using PPO.
r_k

Fig. 1. ChatGPT diagram [2].

Open AI has developed several versions of GPT, which have been compared in Table I, [2] [15] [16].

The development of chatbots has undergone significant evolution throughout the decades. Table II presents a comprehensive overview of chatbots, covering from first robot Eliza in 1966 to the most recently released bot, Bard, in 2023.

TABLE I. COMPARISON OF CHATGPT MODELS

Model	Description	Price	Released	Parameters	Benefits
ChatGPT-1	The initial version of the language model employing transformer architecture	N/A	June 2018	117 million	Text generation, translation, text summarization, language modeling
ChatGPT-2	GPT-1 but modified normalization	N/A	February 2019	1.5 billion	ability to produce realistic text sequences
ChatGPT-3	GPT-3 models can both comprehend and produce natural language. These models were replaced by models of the GPT-3.5 generation, which were more powerful.	Free	June 2020	175 billion parameters	Questions answering, chatbots, and automated content generation
ChatGPT-3.5	A model can comprehend and generate code or natural language. works well for regular task completion.	Free	January 2022	175 billion parameters	cost-effective
ChatGPT-4	A large multimedia model that accepts text inputs and emits text outputs today and will accept image inputs in the future is capable of solving complex problems with greater precision than previous models.	$20	November 2022	1 trillion parameters	Significantly more capable than previous models.

TABLE II. COMPARISON OF CHATBOTS FROM MOST RECENT TO EARLIEST [7,16–18]

Chatbot	Description	Launched Year	Developed by	Available Languages	Parameters
Bard	Is a generative and conversational AI chatbot, that can help with creative tasks, explaining complex topics, and generally distilling information from various sources on the internet. It can also handle nuanced queries.	2023	Google AI	Over 10 languages	137 billion parameters
GPT-4	Is the most recent model in the GPT series; it improves upon GPT-3's strengths and overcomes some of its weaknesses.	2023	OpenAI	Over 17 languages including Arabic	170 trillion parameters
LaMDA	Language Model for Dialogue Applications—LaMDA. It's a text-trained ML model that predicts words and phrases. A human-like chatbot was created.	2021	Google AI	English	137 billion parameters
Tess	Is a mental health chatbot that uses text message conversations to coach people through challenging times.	2020	Google AI	Over 10 languages	N/A
Wysa	Wysa: Anxiety, Therapy Tracker, an AI-powered chatbot, helps people with anxiety, stress, and depression. Wysa helps users modify their negative thoughts and habits using CBT.	2018	Wysa Health Company	English & Hindi	N/A
Replika	The most popular chatbot friend. AI mimics you. Interact with you and gather social media data to learn everything. Replika chatbots listen and take notes like therapists. It replicates you.	2017	Luka company	English & Spanish	137 billion parameters
Woebot	Is an automated conversational agent (chatbot) that aids in self-awareness and mood monitoring. and becomes increasingly tailored to what you need over time.	2016	Woebot Labs Company	English & Spanish	N/A
Google Assistants	Is the next generation of Google Now. It has a more advanced AI, a gentler, more conversational interface, and predicts the information needs of users.	2016	Google	Over 100 languages	137 billion parameters
Amazon Alexa	It has been built into home automation and entertainment devices, making the Internet of Things (IoT) more accessible to people.	2014	Amazon	Over 10 languages	11 billion parameters
Xiaolce	Is Chinese AI. One of the most famous girl chatbots of the era, it was before Tay.	2014	Baidu	English, Japanese, Korean, and Chinese (Traditional &Simplified)	1.07 billion parameters
Microsoft Cortana	It recognizes voice commands, identifies time and place, supports people-based reminders, sends emails and messages, makes and manages lists, chitchats, plays games, and seeks user-requested information.	2014	Microsoft	Over 17 languages including Arabic	10 billion parameters
Mitsuku	Is a chatbot with AI that can tell you stories, quips, and horoscopes. She is able to perform simple games with you and shows you web pages and images from the internet.	2012	Rollo Carpenter, a British Engineer & Software Developer	Over 10 languages	400 million parameters
Apple Siri	Siri prioritizes productivity. No random questions. Siri is a voice-based computer interface, not a person you can chat with.	2011	Apple Inc	20 languages	10 billion parameters
ALICE	ELIZA inspired ALICE, the first internet chatbot. Pattern-matching without actual conversation perception.	1995	Richard Wallace	Wide range of languages	Does not have a fixed number of parameters
PARRY	A chatterbot that simulates human conversation in an amusing, hilarious way.	1972	Kenneth Colby, a psychiatrist at Stanford University	Not support any spoken languages	2,000 parameters
Eliza	First chatbots that can only discuss a specific domain of topics (the point of beginning for chatbots)	1966	Joseph Weizenbaum, a professor of computer science at MIT	10 Languages	No parameters

C. Ethical Concerns

A computer program named Racter was listed as the author of a text in Omni Magazine in November of 1981 [11]. After that in 1984, Racter's book was published as the first book written in a computer program [19]. Since then, due to Racter and AI, copyright issues have received a lot of consideration [20].

Here's the big question: Is it possible to consider a chatbot as an author, given the emergence of AI and chatbots as well as their wide application in various industries, especially in the field of education and student assistance?

The author referred to [21] paper, which stated that using an AI text generator without appropriate attribution would be considered plagiarism. Plagiarism is the unauthorized use of the work or ideas of another individual without citing the source. This action applies to both content generated by human as well as AI, according to the authors in papers [21, 22], when using AI tools, it is essential to cite a source reference. However, significant journals such as Nature and Science have stated that AI chatbots cannot be authors of articles that are published in their journals. This is because the editorial policies on the authorship of these journals state that "AI chatbots do not currently satisfy our authorship criteria" [23].

The reason that AI chatbots cannot be writers is not because they are not human; rather, it is because they do not meet the standards that are currently required. However, in the future, AI chatbots may be accredited as authors of academic articles if they meet the required criteria [11]. Another reason for not being considered as an author is because it is unable to provide permission for these papers to be published, and this is the copyright privacy argument [24].

The use of chatbots to create academic content raises the following ethical concerns:

Bias is defined as a systematic error in decision-making processes that produce unfair results. Bias can originate from a variety of sources, such as data collection, algorithm design, and human interpretation. Machine learning models, a type of artificial intelligence system can learn, and replicate bias patterns found in the training data, which leads to unfair and discriminatory results [25].

Informed consent is a fundamental ethical principle for human subjects' research. It refers to the process of obtaining a participant's consent after providing them with sufficient information about the research, its potential risks, benefits, and alternatives, as well as the opportunity to ask questions and clarify any confusion [26]. On the other hand, there is a lack of consent from AI when researchers use AI-generated text. Therefore, to obtain valid and reliable results, researchers must be aware of the technology they employ in order of transparency and informed consent [22].

Privacy concerns, chatbots generate text from massive datasets, which may contain private and sensitive data. Due to the rising use of chatbot-generated text in scientific research, it is necessary to ensure the privacy of participants' data. Disclosure of this information violates the privacy of the individual and raises significant ethical issues [22].

Research integrity, the information and data retrieved from chatbots such as Bard and ChatGPT may be misleading; in many cases, the chatbot fabricated references that did not exist, such as the article written in this paper [27], using Chat GPT, indeed the references were fake. This confirms that entire reliance on chatbots is unreliable and unsuitable for use and that researchers are responsible for the quality and reliability of their research.

D. Chatbot Researches

The authors of the study [28], predict that ChatGPT will impact every aspect of society. To test ChatGPT, they investigated by writing an academic paper. The results indicate that ChatGPT can assist researchers in writing a cogent (partial) paper that is accurate, informative, and systematic and that the writing is very effective in two to three hours despite the author's limited professional knowledge. Based on user experience, the author considers the potential effects of ChatGPT and other AI tools and concludes with a proposal to modify the learning objectives so that students are taught how to use AI tools, and that education focuses on developing students' creativity and critical thinking, which cannot be replaced by AI tools.

In [22] study, the researcher discussed ethical concerns regarding the use of ChatGPT in scientific research, such as transparency, bias, informed consent, privacy, accountability, and integrity. The researcher concluded that researchers should declare and acknowledge their use of ChatGPT in the research methodology section and stick to research ethics and integrity.

In another investigation on the impact of AI on ethical issues, this time in the context of medical publishing practices in a paper [29], the authors requested ChatGPT contribute a commentary to Lancet Digital Health concerning AI and medical publishing ethics. They also asked ChatGPT how the editorial team can handle the AI-generated academic content. According to ChatGPT's response to their question, Lancet Digital Health should "carefully consider the ethical implications of publishing articles produced by AI."

In an additional paper [30], the author posed the following question to ChatGPT: "When streptozotocin-induced diabetes is prepared in growing rats, can you predict its effect on the facial bone growth pattern?" ChatGPT responded to the inquiry. Then he inquired about the references for this topic, and ChatGPT attached all the references along with their authors. Then, the researcher investigated whether these references were real or fake, and found out that they were all fake. However, the researcher confirms that ChatGPT did an excellent job editing English grammar. The researcher concluded that any novel ideas generated by ChatGPT must be validated, that results must be verified before publication, and lastly, any AI assistance must be disclosed.

In continuance of the preceding article debating whether ChatGPT references are real or fake. Through ChatGPT, the author of the paper [27], completes the article without human intervention or editing. Unfortunately, upon investigation, the references that the author requested from ChatGPT were found to be fake.

The author of the paper [31], discusses a variety of ChatGPT-related topics, including its history, applications, challenges, bias, ethics, limitations, and future. In terms of ethics, topics discussed include data privacy, bias, transparency, autonomy, human agency, emotional manipulation, persuasion, and reliance on AI-generated content, as well as others. It was concluded that ChatGPT made significant contributions to scholarly research in terms of linguistically coherent text generation, is grammatically accurate, and has the potential to transform the field in the future if its challenges and ethical issues are addressed.

In addition to the features offered by chatbots, there are also cyber risks that must be addressed. The author of the paper [14], investigated the cyber risks associated with the use of ChatGPT and other chatbots based on AI that is similar to ChatGPT. Also, the vulnerabilities that are exploited by malicious actors because ChatGPT risks providing simple scripting and access to coding by cybercriminals, as well as ways to mitigate them, such as limiting entry barriers for cybercriminals, complying with regulations, and more.

A preliminary study was conducted by the authors of the [32] paper, comparing ChatGPT translations to those created by human translators. They found that although the translations were not always accurate, they performed competitively with commercial translation products, such as Google in high-resource European languages but significantly different in low-resource languages.

The authors of the paper [33], evaluated ChatGPT's performance on the United States Medical Licensing Examination (USMLE), which consists of three examinations and does not involve any specialized training or reinforcement. They identified two major themes: (1) the increasing accuracy of ChatGPT approaching or exceeding the threshold for passing the USMLE, and (2) the potential for AI to generate new insights that could aid human learners in the context of medical education.

Furthermore, discussing ethical issues related to technology use in academia, scholarly research, and publishing, the paper's authors in [34], compare the effect of GPT/ChatGPT versus that of other language paradigms. GPT3 has proven to be flexible, efficient, and capable of generating human-like language, which makes it useful for tasks such as translating, annotating, and answering questions. Additionally, ChatGPT has the potential to enhance search efficiency and the quality of academic publications. On the one hand, they discussed the ethical considerations that need to be taken into account, such as the ownership of the content, since it is not clear who owns the rights to the generated text and the copyright issues that are of concern yet.

III. Materials and Methods

The present research methodology involves two phases, the first phase is a qualitative method which includes: experimentation with two types of chatbots to examine and assess the content generated by chatbots.

Chatbots is a computer program that imitates dialogue with human users, typically employing natural language processing (NLP) to analyze inputs and generative AI to automatically generate responses. Chatbots will go through experiments: ChatGPT-4 which is an artificial intelligence-based chatbot, was developed by OpenAI. It was built on top of OpenAI's GPT-3.5 and GPT-4 families of LLMs [2]. In addition, Bard is a chatbot. It is also based on LLM powered by Google. Improved and lightweight of LaMDA [3].

Initially, the ChatGPT-4 and Bard will go through experiments by a request for five academic articles related to the field of technology, all having a word count of 1000. The credibility of references will be manually verified, and the percentage of plagiarized content will be measured. Moreover, six articles were generated by ChatGPT-4, to increase the dataset in various fields with a count of 500 words.

The second phase is a quantitative method that involves applying topic modeling, which is an unsupervised machine-learning approach that involves scanning numerous documents, articles, feedback, or emails. In this study, the focus will be on articles produced by chatbots. The primary objective of this technique is to identify patterns of words and phrases within the articles, regardless of their relevance to the topic. to its underlying semantic structure. The given entity can be described as a collection of words without any specific order or structure [35].

Then, the analysis using Latent Dirichlet Allocation (LDA) is a computational model that shares similarities with Latent Semantic Analysis (LSA). However, LDA differs from LSA in that it assigns topics to word order, with the aim of ascertaining the composition of topics in documents [35].

In the end, an assessment will be conducted to evaluate the accuracy of the content generated by chatbots based on the bibliography. Fig. 2 demonstrates the phases, and given in detail.

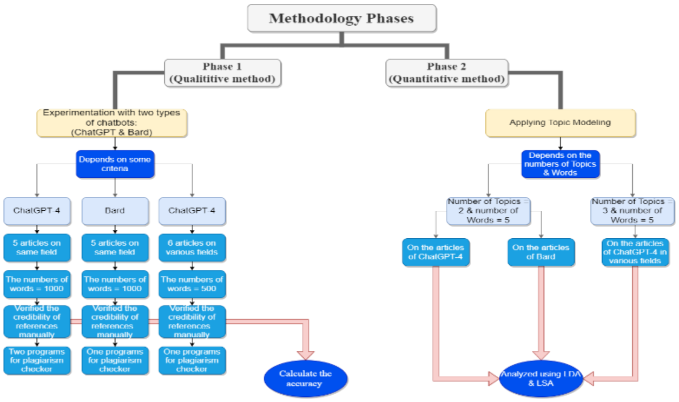

Methodology Phases
Phase 1 (Qualititive method)
Phase 2 (Quantitative method)
Experimentation with two types of chatbots: (ChatGPT & Bard)
Depends on some criteria
ChatGPT 4
Bard
ChatGPT 4
5 articles on same field
5 articles on same field
6 articles on various fields
The numbers of words = 1000
The numbers of words = 1000
The numbers of words = 500
Verified the credibility of references manually
Verified the credibility of references manually
Verified the credibility of references manually
Two programs for plagiarism checker
One programs for plagiarism checker
One programs for plagiarism checker
Calculate the accuracy
Applying Topic Modeling
Depends on the numbers of Topics & Words
Number of Topics 2 & number of Words = 5
Number of Topics 3 & number of Words = 5
On the articles of ChatGPT-4
On the articles of Bard
On the articles of ChatGPT-4 in various fields
Analyzed using LDA & LSA

Fig. 2. Methodology phases.

IV. Results

This section presents a comparative analysis of the chatbots ChatGPT-4 and Bard, as well as a topic modeling approach applied to the articles stated in Tables III, IV, and V which were analyzed afterward.

A. *Comparison of the Academic Articles Generated by GPT-4 in the Various Fields*

Academic articles generated by GPT-4 in a variety of scientific disciplines, such as technology, medicine, space, earth, society, and marketing. To assess the breadth of his knowledge and reliability of his sources in various fields. Table III compares based on the same word count, the number of bibliographies, and the Plagiarism ratio.

B. Comparison of the Academic Articles Generated by GPT-4 in the Same Fields

Academic articles generated by GPT-4 in the same scientific disciplines, such as technology. To assess the accuracy of his knowledge and reliability of his sources in the same fields. Table IV compares based on the same word count, the number of bibliographies, and the Plagiarism ratio in the two programs.

C. Comparison of the Academic Articles Generated by Bard in the Same Fields

Academic articles generated by Bard in the same scientific disciplines, such as technology. To assess the accuracy of his knowledge and reliability of his sources in the same fields to compare it with ChatGPT-4. Table V compares based on the same word count, the number of bibliographies, and the Plagiarism ratio.

The accuracy will be computed after performing a comparison based on specified criteria presented in Tables III, IV, and V as the following equation:

$$Accuracy = \frac{corrected\ bibliographies}{The\ total\ number\ of\ bibliographies} \qquad (1)$$

The accuracy of ChatGPT-4-generated articles in Table III is 77.5%, while in Table IV, it's 96%. Bard's articles in Table V have an accuracy of 52%. These results are visually summarized in Fig. 3, which provides a comprehensive overview of the findings from Tables III, IV, and V.

D. Topic Modeling

Topic modeling has been applied to articles in Table IV. The outputs of LDA and LSA for two topics and five words for ChatGPT-4 articles are shown in Tables VI and VII respectively. The results are visualized in Fig. 4 and Fig. 5.

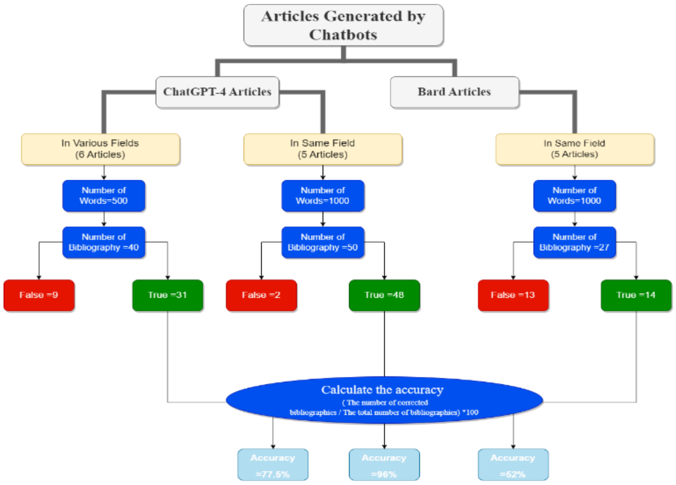

Articles Generated by Chatbots
ChatGPT-4 Articles
Bard Articles
In Various Fields (6 Articles)
In Same Field (5 Articles)
In Same Field (5 Articles)
Number of Words=500
Number of Words=1000
Number of Words=1000
Number of Bibliography =40
Number of Bibliography =50
Number of Bibliography =27
False =9
True =31
False =2
True =48
False =13
True =14
Calculate the accuracy
(The number of corrected bibliographies / The total number of bibliographies) *100
Accuracy =77.5%
Accuracy =96%
Accuracy =52%

Fig. 3. Diagram for findings (Based on Tables III, IV, and V).

TABLE III. COMPARISON OF THE ACADEMIC ARTICLES GENERATED BY GPT-4 IN THE VARIOUS FIELDS

Article Name	Field	Number of words	The actual word without a Bibliography	Number of Bibliography	True Bibliography	False Bibliography	Plagiarism Checker X
1. Artificial Intelligence in Medical Health: Transforming Healthcare Through Advanced Technologies	Technology & Medicine	500	510	6	6	0	15%
2.The Impact of social media on E-commerce: Exploring the Interplay of Online Interactions and Commerce	Society & Marketing	500	520	5	4	1 in publication year	22%
3.Harnessing the Power of IoT in Smart Cities: Opportunities and Challenges	Technology	500	547	7	4	32 in publication year 1 in both years and the author's name	18%
4.The Impact of Fake Hashtags on Twitter: Unraveling the Consequences of Misleading Trends and Manipulated Discourse	Society & Technology	500	574	9	7	2 in publication year	10%
5.The Impact of Online Advertising on Purchases: A Multi-faceted Examination of Consumer Responses to Digital Marketing Efforts	Society & Marketing &Technology	500	551	6	6	0	14%
6.The Possibility of Life Beyond Earth: Investigating Extraterrestrial Habitats and Astro Biological Discoveries	Space & Earth	500	577	7	4	31 all the paper 2 in publication year	21%

TABLE V. COMPARISON OF THE ACADEMIC ARTICLES GENERATED BY GPT-4 IN THE SAME FIELDS

Article Name	Field	Number of words	The actual word without a Bibliography	Number of Bibliography	True Bibliography	False Bibliography	Plagiarism Checker X	Plagiarism Checker in iThunticate
1. Machine Learning for Financial Forecasting: Techniques, Applications, and Challenges	Technology	1000	711	9	9	0	24%	25%
2. Machine Learning for Image and Video Processing: Techniques, Applications, and Challenges	Technology	1000	686	9	9	0	30%	22%
3. Machine Learning for Medical Diagnosis: Current Advances and Future Perspectives	Technology	1000	821	10	10	0	19%	18%
4. Machine Learning for Natural Language Processing: A Comprehensive Overview	Technology	1000	610	10	10	0	30%	25%
5. Machine Learning for Social Media Analysis: A Comprehensive Overview	Technology	1000	873	12	10	2 in the author's name & publication year	29%	25%

TABLE VI. COMPARISON OF THE ACADEMIC ARTICLES GENERATED BY BARD IN THE SAME FIELDS

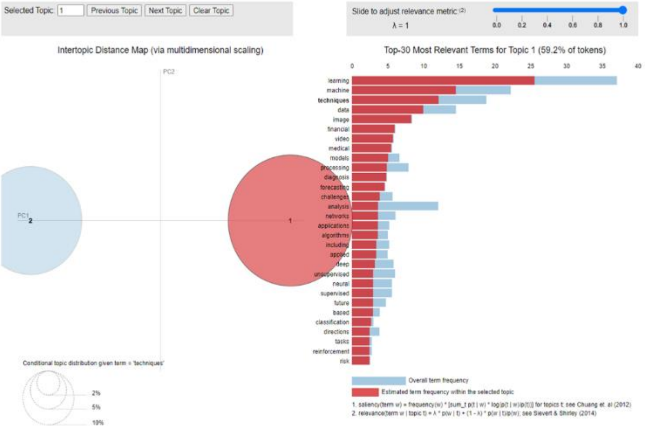
Article Name	Field	Number of words	The actual word without a Bibliography	Number of Bibliography	True Bibliography	False Bibliography	Plagiarism Checker X
1. Artificial Intelligence in Medical Health: Transforming Healthcare Through Advanced Technologies	Technology	1000	545	3	0	32 papers not found.1 in title	33%
2. The Impact of Social Media on E-commerce: Exploring the Interplay of Online Interactions and Commerce	Technology	1000	599	5	2	31 in title not found.1 in the author's name & publication year 1 in title not found & publication year	25%
3. Harnessing the Power of IoT in Smart Cities: Opportunities and Challenges	Technology	1000	522	6	4	21 Found 2 papers different in authors or title 1 paper not found	28%
4. The Impact of Fake Hashtags on Twitter: Unraveling the Consequences of Misleading Trends and Manipulated Discourse	Technology	1000	562	6	5	1 paper not found	28%
5. The Impact of Online Advertising on Purchases: A Multi-faceted Examination of Consumer Responses to Digital Marketing Efforts	Technology	1000	499	7	3	42 papers not found 1 in the title name 1 in publication year	34%

TABLE VII. LDA MOST RELEVANT TERMS FOR TWO TOPICS AND FIVE WORDS FOR CHATGPT-4

Topic Number	Terms	Rate
1	Learning	0.054
1	Machine	0.031
1	Techniques	0.026
1	Data	0.021
1	Image	0.018
2	Learning	0.035
2	Analysis	0.026
2	Machine	0.024
2	Social	0.021
2	Techniques	0.020

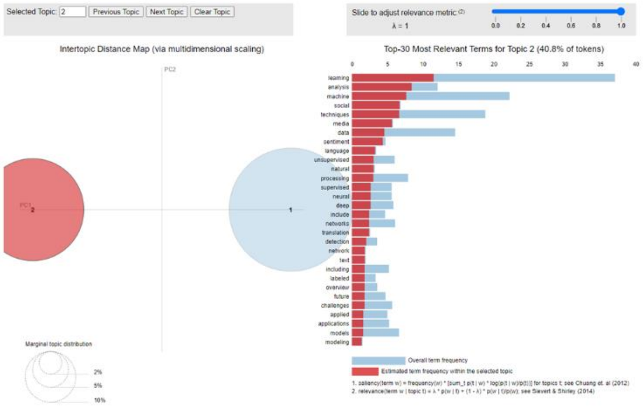
TABLE VIII. LSA MOST RELEVANT TERMS FOR TWO TOPICS AND FIVE WORDS FOR CHATGPT-4

Topic Number	Terms	Rate
1	Learning	- 0.603
1	Machine	-0.356
1	Techniques	-0.312
1	Data	-0.236
1	Analysis	-0.197
2	Social	0.480
2	Analysis	0.414
2	Media	0.400
2	Image	-0.222
2	Medical	-0.187



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Fig. 4. Topic modeling results for topic 1 with five words for ChatGPT-4 articles.



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Fig. 5. Topic modeling results for topic 2 with five words for ChatGPT-4 articles.

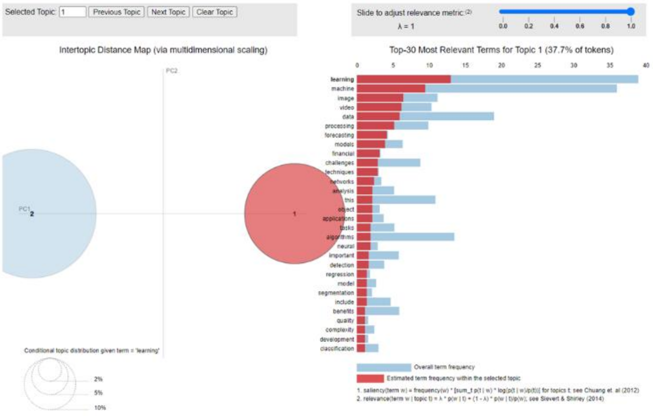
For the article in Table V, the output of LDA and LSA for two topics and five words for Bard articles are shown in Tables VIII and IX and visualized in Fig. 6 and Fig. 7.

TABLE IX. LDA MOST RELEVANT TERMS FOR TWO TOPICS AND FIVE WORDS FOR BARD ARTICLES

Topic Number	Terms	Rate
1	Learning	0.046
1	Machine	0.034
1	Image	0.023
1	Video	0.022
1	Data	0.021
2	Machine	0.057
2	Learning	0.056
2	Data	0.028
2	Algorithms	0.025
2	This	0.019

TABLE X. LSA MOST RELEVANT TERMS FOR TWO TOPICS AND FIVE WORDS FOR BARD ARTICLES

Topic Number	Terms	Rate
1	Learning	0.571
1	Machine	0.489
1	Data	0.269
1	Image	0.209
1	Video	0.196
2	Image	-0.358
2	Video	-0.344
2	Processing	-0.262
2	Machine	0.258
2	algorithms	0.246



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Fig. 6. Topic modeling results for topic 1 with five words for bard articles.

Topic modeling was also applied in combining the articles in Table III and Table IV using three topics. The output of LDA and LSA for three topics and five words based on ten articles are shown in Tables X and XI and visualized in Fig. 8, Fig. 9, and Fig. 10.

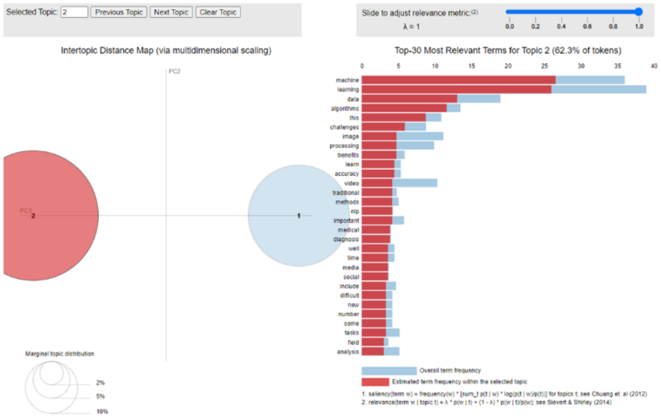

Selected Topic: 2
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 2 (62.3% of tokens)
0
5
10
15
20
25
30
35
40
machine
learning
data
algorithms
this
challenges
image
processing
benefits
learn
accuracy
video
traditional
methods
nlp
important
medical
diagnosis
well
time
media
social
include
difficult
new
number
some
tasks
field
analysis
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Fig. 7. Topic modeling results for Topic 2 with five words for bard articles.

TABLE XI. LDA MOST RELEVANT TERMS FOR THREE TOPICS AND FIVE WORDS BASED ON TEN ARTICLES

Topic Number	Terms	Rate
1	IoT	0.014
1	Media	0.012
1	Social	0.012
1	Data	0.012
1	AI	0.012
2	Advertising	0.030
2	online	0.019
2	Ads	0.017
2	Consumer	0.009
2	Users	0.009
3	Learning	0.043
3	Machine	0.025
3	Techniques	0.022
3	Data	0.016
3	Analysis	0.014

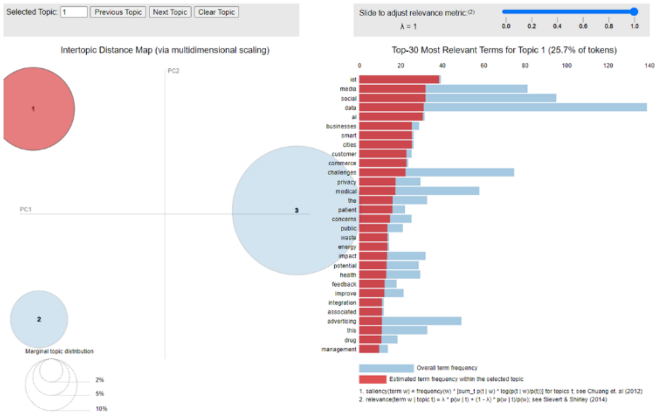

Selected Topic: 1
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 1 (25.7% of tokens)
0
20
40
60
80
100
120
140
iot
media
social
data
ai
businesses
smart
cities
customer
commerce
challenges
privacy
medical
the
patient
concerns
public
waste
energy
impact
potential
health
feedback
improve
integration
associated
advertising
this
drug
management
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Fig. 8. Topic modeling results for Topic 1 with five words for ten articles.

TABLE XII. LSA MOST RELEVANT TERMS FOR THREE TOPICS AND FIVE WORDS BASED ON TEN ARTICLES

Topic Number	Terms	Rate
1	IoT	0.014
1	Media	0.012
1	Social	0.012
1	Data	0.012
1	AI	0.012
2	Advertising	0.030
2	online	0.019
2	Ads	0.017
2	Consumer	0.009
2	Users	0.009
3	Advertising	0.403
3	Hashtags	-0.342
3	Fake	-0.291
3	Online	0.260
3	Social	-0.248

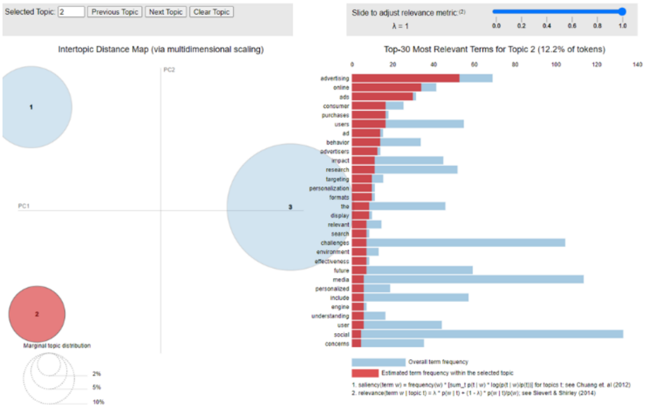

Selected Topic: 2
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 2 (12.2% of tokens)
0
20
40
60
80
100
120
140
advertising
online
ads
consumer
purchases
users
ad
behavior
advertisers
impact
research
targeting
personalization
formats
the
display
relevant
search
challenges
environment
effectiveness
future
media
personalized
include
engine
understanding
user
social
concerns
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Fig. 9. Topic modeling results for Topic 2 with five words for ten articles.

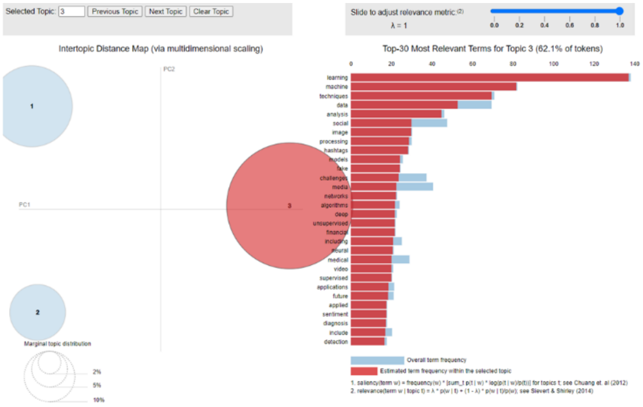

Selected Topic: 3
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 3 (62.1% of tokens)
0
20
40
60
80
100
120
140
learning
machine
techniques
data
analysis
social
image
processing
hashtags
models
fake
challenges
media
networks
algorithms
deep
unsupervised
financial
including
neural
medical
video
supervised
applications
future
applied
sentiment
diagnosis
include
detection
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Fig. 10. Topic modeling results for Topic 3 with five words for ten articles.

V. Discussion

Based on the experiences of Bard and ChatGPT-4, the latter has not exceeded a plagiarism rate of 30 percent, whereas Bard did exceed it. The word count of ChatGPT-4 exceeds Bard, although it falls short of the prescribed 1000-word limit. Four out of five references for ChatGPT-4 were found to be real.

The references were found to be entirely accurate but with some errors. Particularly, the fifth article in Table IV contained errors in the names of the authors and the year of publication. On the other hand, it is notable that no article was found in Bard to have completely correct references. Moreover, accessing the complete scientific papers demonstrated challenges for Articles 1, 3, 4, and 5, as they were not easily found. This made it challenging to verify the authenticity of these papers. The author's names and the title appear to be incorrect and vice versa.

According to Table III, the findings of ChatGPT-4 were comparatively accurate in terms of word count, as all the generated texts were within the requested limit of 500 words, with an additional few words. This comparison was conducted across different fields. There are errors in publication year and author names, but 1 existence of a particular paper is not found. Upon inquiry regarding the ChatGPT-4 fake reference, the citation was altered to a more appropriate and true reference. Furthermore, the percentage of plagiarism did not exceed 25%.

After applying topic modeling to ChatGPT-4 articles, determine whether the article's content is related to its title. The LDA model demonstrated that the most relevant word was (learning) with a percentage of 0.054 for Topic 1 and (learning) for Topic 2 with 0.035, compared to the LSA model indicated that the most relevant word was (learning) with a percentage of - 0.0603 for Topic 1 and (social) for Topic 2 with 0.480. Tables VI and VII display every word of LDA and LSA. To visualize the output of the LDA model, refer to Fig 4. 59.2% of the tokens are associated with Topic 1. In Fig. 5. 40.8% of the tokens are associated with Topic 2.

In Bard, the most relevant word in the LDA model to Topic 1 was (learning) with a percentage of 0.046 for Topic 1, and (machine) with a percentage of 0.057 for Topic 2. Compared with LSA the most relevant word to Topic 1 is (learning) with a percentage of 0.571 and -0.358 for the word (image) in Topic 2. Tables VIII and IX list all the words LDA and LSA. To visualize the LDA, Fig. 6 showed 37.7% of tokens for the most relevant words for Topic 1, and Fig. 7, showed 62.3% of tokens for Topic 2.

Tables X and XI showed a bigger set of articles and increased the Topics to 3 and in different fields a combination of articles from Tables III and IV. The most relevant words for LDA are (IoT) with a percentage of 0.014 for Topic 1 as shown in Fig. 8, with 25.7% of tokens relevant words. And (advertising) with a percentage of 0.030 for Topic 2 as shown in Fig. 9, with 12.2% of tokens relevant words. Lastly (learning) with a percentage of 0.043 for Topic 3 as shown in Fig. 10, with 62.1% of tokens relevant words.

The ChatGPT-4 bibliography exhibits a greater degree of accuracy in comparison to the Bard bibliography as shown in equation 1. However, it is important to note that the accuracy of the ChatGPT-4 bibliography is not absolute, and there may be instances where it is erroneous, given that several bibliographies were found to be incorrect which agreed with the previous papers [27] and [30], that found the bibliographies are fake and must be validated before use it. Students and researchers should not rely on chatbots because their accuracy is insufficient for 100% error-free citations, and each researcher is responsible for the credibility and accuracy of his research and information.

Due to its lack of absolute accuracy chatbots, it cannot be considered an author of scientific papers. In addition to credibility and integrity, scientific research depends on humans, and a chatbot cannot be regarded as human. It can be used for time-consuming routine tasks that align with previous papers [34], but in academic research, it lacks direct information and alters its responses if you disagree with it. In contrast to chatbots, the researcher understands all aspects of his research topic and can defend his opinions, but the researcher's ethics and integrity will determine the achievement of his research.

VI. Conclusion

The advancement of chatbots is at an interesting growth, and the use of chatbots in higher education can yield numerous benefits such as text summarization, text generation, and translation. However, the utilization of these chatbots also poses several challenges and concerns, particularly concerning the context of academic integrity and the problem of plagiarism. The utilization of these chatbots may potentially boost fraud and distinguishing between automated and human-generated writing could be a challenge.

According to the findings of the present analysis, the utilization of two chatbots (ChatGBT-4 and Bard), by applying topic modeling even if it is shown relevant to the topic but cannot be considered scholarly or suitable for academic research, as several criteria assessed in the present investigation were not achieved.

Most of the articles generated by Bard were found to be inaccurate and some non-existent. Although ChatGBT-4 exhibits greater accuracy than Bard, it remains insufficient, failing to achieve the level of accuracy characteristic of human-generated information. Because of the change in responses depending on the conversation context.

Furthermore, there exist numerous ethical standards concerning copyright and research ethics. Chatbots must not be considered authors because of their inability to provide accurate information. Higher education institutions should carefully consider the potential risks that chatbots pose to students and they must develop a well-planned strategy for educating and informing students about the use policy of chatbots. In addition, developing a set of tools that detect plagiarism and protect academic ethics and integrity.

In future investigations, the assessment of scientific articles in ChatGBT-4 will be expanded with the incorporation of updated information more than the year 2021. Similarly, the accuracy of Bard will be further evaluated once it progresses beyond the experimental phase.

REFERENCES

[1] M. Rahaman, M. M. Ahsan, N. Anjum, M. Rahman, and M. N. Rahman, “The AI Race is On! Google’s Bard and OpenAI’s ChatGPT Head to Head: An Opinion Article,” Mizanur and Rahman, Md Nafizur, The AI Race is on, 2023.

[2] “Introducing ChatGPT,” OpenAI, Nov. 30, 2022. https://openai.com/blog/chatgpt#OpenAI (accessed Apr. 28, 2023).

[3] S. ’Hsiao and E. ’Collins, “AI Try Bard and share your feedback,” Google, Mar. 21, 2023. https://blog.google/technology/ai/try-bard/ (accessed May 16, 2023).

[4] Y. ’ ’Mehdi, “Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web,” Microsoft, Feb. 07, 2023. https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/ (accessed May 16, 2023).

[5] A. M. Turing, Computing machinery and intelligence. Springer, 2009.

[6] “Jabberwacky,” Nov. 15, 2019. https://en.wikipedia.org/wiki/Jabberwacky (accessed Apr. 30, 2023).

[7] E. Adamopoulou and L. Moussiades, “Chatbots: History, technology, and applications,” Machine Learning with Applications, vol. 2, p. 100006, 2020.

[8] H. Bansal and R. Khan, “A review paper on human computer interaction,” Int. J. Adv. Res. Comput. Sci. Softw. Eng, vol. 8, no. 4, p. 53, 2018.

[9] A. Khanna, B. Pandey, K. Vashishta, K. Kalia, B. Pradeepkumar, and T. Das, “A study of today’s AI through chatbots and rediscovery of machine intelligence,” International Journal of u-and e-Service, Science and Technology, vol. 8, no. 7, pp. 277–284, 2015.

[10] B. A. Shawar and E. Atwell, “Chatbots: are they really useful?,” Journal for Language Technology and Computational Linguistics, vol. 22, no. 1, pp. 29–49, 2007.

[11] J. Y. Lee and S. Huh, “Can an artificial intelligence chatbot be the author of a scholarly article?,” jeehp, vol. 20, no. 0, p. 6, Feb. 2023, doi: 10.3352/jeehp.2023.20.6.

[12] A. Vaswani et al., “Attention is all you need,” Adv Neural Inf Process Syst, vol. 30, 2017.

[13] M. Ruby, “How ChatGPT Works: The Model Behind The Bot,” towardsdatascience, Jan. 30, 2023. https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286 (accessed Apr. 29, 2023).

[14] G. Sebastian, “Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk?,” International Journal of Security and Privacy in Pervasive Computing, vol. 15, no. 1, pp. 1–11, Mar. 2023, doi: 10.4018/IJSPPC.320225.

[15] T. ’Desk, “Open AI’s GPT 4 could support up to 1 trillion parameters, will be bigger than ChatGPT 3,” The Indian EXPRESS - JOURNALISM OF COURAGE, Jan. 23, 2023. https://indianexpress.com/article/technology/tech-news-technology/chatgpt-4-release-features-specifications-parameters-8344149/ (accessed Apr. 30, 2023).

[16] F. ’Ali, “GPT-1 to GPT-4: Each of OpenAI’s GPT Models Explained and Compared,” MAKE USE OF, Apr. 11, 2023. https://www.makeuseof.com/gpt-models-explained-and-compared/ (accessed Apr. 30, 2023).

[17] C. ’Wankhede, “What is Google’s Bard AI? Here’s everything you need to know,” ANDROID AUTHORITY, Mar. 22, 2023. https://www.androidauthority.com/google-bard-chatbot-3295464/ (accessed May 29, 2023).

[18] K. ’Rajnerowicz, “26 Best Real Life Chatbot Examples [Well-Known Brands],” TIDIO, Apr. 07, 2023. https://www.tidio.com/blog/chatbot-examples/#siri (accessed May 29, 2023).

[19] “The First Book Written by a Computer Program,” “Normans,Jeremy,” 1984. https://www.historyofinformation.com/detail.php?id=3351 (accessed May 01, 2023).

[20] T. L. Butler, “Can a computer be an author-copyright aspects of artificial intelligence,” Comm/Ent LS, vol. 4, p. 707, 1981.

[21] B. L. Frye, “Should Using an AI Text Generator to Produce Academic Writing Be Plagiarism?,” Fordham Intellectual Property, Media & Entertainment Law Journal, Forthcoming, 2022.

[22] Z. N. Khlaif, “Ethical Concerns about Using AI-Generated Text in Scientific Research,” Available at SSRN 4387984, 2023.

[23] Springer Nature, “Authorship,” Springer Nature, 2023. https://www.nature.com/nature-portfolio/editorial-policies/authorship (accessed May 01, 2023).

[24] C. Stokel-Walker, “ChatGPT listed as author on research papers: many scientists disapprove,” Nature.

[25] E. Ferrara, “Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, And Mitigation Strategies,” arXiv preprint arXiv:2304.07683, 2023.

[26] S. ’ ’Manti and ’Licari.Amelia ’, “How to obtain informed consent for research,” National Library of Medicine, Jun. 2018. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5980471/#C2 (accessed May 15, 2023).

[27] M. R. King, “A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education,” Cellular and Molecular Bioengineering, vol. 16, no. 1. Springer, pp. 1–2, Feb. 01, 2023. doi: 10.1007/s12195-022-00754-8.

[28] X. Zhai, “ChatGPT user experience: Implications for education,” Available at SSRN 4312418, 2022.

[29] M. Liebrenz, R. Schleifer, A. Buadze, D. Bhugra, and A. Smith, “Generating scholarly content with ChatGPT: ethical challenges for medical publishing,” Lancet Digit Health, vol. 5, no. 3, pp. e105–e106, 2023.

[30] S.-G. Kim, “Using ChatGPT for language editing in scientific articles,” Maxillofac Plast Reconstr Surg, vol. 45, no. 1, p. 13, 2023.

[31] P. P. Ray, “ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope,” Internet of Things and Cyber-Physical Systems, 2023.

[32] W. Jiao, W. Wang, J. Huang, X. Wang, and Z. Tu, “Is ChatGPT a good translator? A preliminary study,” arXiv preprint arXiv:2301.08745, 2023.

[33] T. H. Kung et al., “Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models,” PLoS digital health, vol. 2, no. 2, p. e0000198, 2023.

[34] B. Lund, T. Wang, N. R. Mannuru, B. Nie, S. Shimray, and Z. Wang, “ChatGPT and a New Academic Reality: AI-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing,” arXiv preprint arXiv:2303.13367, 2023.

[35] ’Pascual.Federico ’, “Topic Modeling: An Introduction,” MonkeyLearn, Sep. 26, 2019. https://monkeylearn.com/blog/introduction-to-topic-modeling/ (accessed May 23, 2023).

A Comparative Study of Stemming Techniques on the Malay Text

Rosmayati Mohemad, Nazratul Naziah Mohd Muhait, Noor Maizura Mohamad Noor, Nur Fadilla Akma Mamat
Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu,
21030 Kuala Nerus, Terengganu, Malaysia

***Abstract*—Text stemming, an essential preprocessing step in the development of Natural Language Processing (NLP) applications, involves the transformation of various word forms into their root words. Stemming plays a critical role in decreasing the volume of text, thereby enhancing the efficiency of various computational tasks such as information retrieval, text classification, and text clustering. Stemming is a rule-based approach. On the other hand, it frequently suffers affixation errors that result in under-stemming, over-stemming, or both, as well as unstemmed or spelling exceptions. Every language has different stemming techniques, and among the most well-known Malay stemming algorithms are the Othman and Ahmad algorithms. Therefore, this study aims to compare the performance of the stemming errors between the Othman and Ahmad algorithms in stemming Malay text, particularly on two different domains of textual datasets, which are the course summaries of the education domain and housebreaking crime reports of the crime domain. The Othman algorithm presents a set of 121 stemming rules (set A). In the meantime, Ahmad's algorithm proposes two distinct sets of stemming rules, comprising 432 (set B) and 561 rules (set C), respectively. Based on the experiment results with 100 course summaries, the Ahmad algorithm (Set B) obtained a higher accuracy rate of 93.61%. The second highest is the Ahmad algorithm (Set C) with 93.53%. The Othman algorithm achieved the lowest accuracy with 86.04% compared to the other two algorithms. Meanwhile, findings from the experiment with 100 housebreaking crime reports show similar results, with the Ahmad algorithm (Set C) achieving the highest stemming accuracy of approximately 93.80% and the Othman algorithm producing the lowest stemming accuracy (83.09%). The result indicates that stemming accuracy is consistent across different types of datasets.**

***Keywords*—Algorithm; ahmad algorithm; malay language; othman algorithm; rule-based; stemming; stemmer**

I. INTRODUCTION

Due to the explosive growth of textual information that has continuously been generated from electronic media over the past ten years, text analytics has received a lot of attention and research [1], [2]. Text analytics, sometimes called text mining, is the integration of linguistic, computational statistics, and computer science techniques [3]. The research on text analytics has been conducted in a variety of languages, including English [4], Arabic [5], Russian [6], Indian [7], Chinese [8], and Thailand [9] to solve various problems in text classification, text clustering, sentiment analysis, and topic modeling. Text analytics have a significant impact on data science [10], [11]. The significant contribution of text analytics in terms of providing enriched and structured datasets that enable in-depth analysis. This has led researchers to explore various text mining techniques, which involve the extraction of valuable information and insights from unstructured textual data. Unstructured textual information requires preprocessing, involving the removal of irrelevant terms, before it can be used in text analytics tasks. Preprocessing is important for decreasing data sparsity, reducing data dimensionality, increasing the quantity of captured semantic information, and ensuring data consistency [12]. The common steps in text preprocessing are tokenization (splitting text into words or phrases), stop word removal, and stemming [13], [14]. Working with a large volume of dimensional text could negatively affect the performance of text analytics. Therefore, stemming is critical for improving the effectiveness of text analytics.

Stemming is one of the basic and essential steps in text preprocessing. It is a natural language processing (NLP) technique used to match the numerous inflectional and derivational morphological forms of a word to its stem or root word. For example, stemming matches the words *maintaining*, *maintained*, and *maintenance* to their root word, *maintain*. Meanwhile, stemmers are the programs that do stemming [15]. In Malay, there are seven-word patterns: affixation, reduplication, compounding, blending, clipping, abbreviation, and borrowing [16]. Malay morphology is recognised for having extremely complex morphological features that are used to construct different word patterns. For instance, the addition of the prefix "pe" to the root word "makan" (eat) results in the word "pemakan" (eater), thereby modifying the meaning of the root word. Thus, it is crucial to understand the morphological structure of the Malay language to reduce derived words into their respective root words.

NLP employs various stemming approaches such as rule-based, dictionary-based, statistical-based, and hybrid stemming. The best approach is determined by the language involved and the nature of the textual dataset. The stemming strategy of rule-based affix elimination is used in this paper, which eliminates the prefix, suffix, and circumfix infix. Othman [17] and Ahmad [18] algorithms are the most pioneering rule-based Malay stemmers. Even though there are plenty of rule-based stemming approaches for Malay that have been improved by the previous researchers since then, they still suffer from affixation errors, including over-stemming, under-stemming, unchanged, and spelling exceptions [19], [20], [21]. The major causes of this stemming error are the affix removal method, the similarity of the root word with the affixation word, and exception rules in prefixation and confixation [22],[23].

The quality of stemming algorithms is typically measured by how accurately they map the variant forms of a word to the same stem. In addition, the presence of diverse morphological structures within a textual dataset, the proper use of appropriate vocabularies, and the word patterns in the textual datasets also play a significant role in contributing to the accuracy of stemming. To the best of our knowledge, there is limited study on the comparative analysis of stemming algorithms for Malay. The most recent works were published in [24] and [25]. However, the purpose of this paper is not to discuss any improvements in terms of the morphological rules of Malays languages. Nevertheless, the purpose of this research is to determine the degree to which the abundance and consistency of precise vocabulary and grammar usage in the textual dataset influence the efficacy of the stemming procedure. Therefore, the objective of this study is to conduct a comparative analysis of the Othman and Ahmad algorithms in terms of their effectiveness in stemming textual data from two distinct domains: education and crime. The analysis of the error rate in stemming and the accuracy of performance for different textual datasets are performed. The analysis is conducted utilising the 121 rules of the Othman algorithm, referred to as set A, as well as the Ahmad algorithm for both set B (comprising 432 rules) and set C (comprising 561 rules). A total of 100 Malay documents for each domain is randomly selected in order to assess the performance of this study. The best result was obtained by the Ahmad algorithm for both datasets, with a stemming accuracy of 93.61% for the education dataset and 93.80% for the crime dataset. Meanwhile, the Othman algorithm attains a stemming accuracy of 86.04% for the education dataset and 83.09% for the crime dataset.

This paper is organised into five sections. Section II discusses the related works on the existing Malay word stemmer. Meanwhile, Section III describes the research methodology used to compare Malay stemming algorithms. Section IV presents the experiment's results and discussion. Finally, Section V concludes the paper by summarising the main achievements and making future recommendations.

II. Related Work

This section discusses a selection of recent studies on Malay stemming algorithms that have employed a rule-based affix elimination approach, which involves the removal of prefixes, suffixes, circumfixes, and infixes. A prefix is a type of affix that is attached to the beginning of a root word, while a suffix is a specific type of affix that is appended to the last position of a root word. In the realm of linguistics, it is worth noting that a linguistic element known as a circumfix, or more formally referred to as a prefix-suffix, is an affix that is comprised of two parts. Both of these parts are strategically positioned, with one located at the beginning of the root word, while the other is attached to the end of the root word. An infix is a type of affix that is inserted in the middle of a root word.

The most pioneering stemmer for Malay is the Othman algorithm, which was proposed in 1993 [22]. This algorithm makes use of Kamus Dewan 1991, a Malaysian dictionary. The use of the dictionary facilitates the identification of the root word after the removal of the affixes, provided that the affixes are matched to the stemming rules. By using the pattern matching rules, the affixes of the word are eliminated once the rule is matched. However, this stemmer has caused over-stemming errors because it does not consider searching for the word in the dictionary before performing the stemming process. Following this, Ahmad et al. [23] proposed a modified version of the Othman algorithm, known as the Ahmad algorithm, in which the algorithm considers dictionary lookups before proceeding to the stemming process and improves the order of applied morphological rules. In addition to enhancing stemming performance, two sets of new rules are developed, each consisting of 432 rules and 561 rules. This algorithm, which uses the rule application order, has been performed on two datasets of ten chapters of the Quran and 10 research abstracts. A series of empirical experiments are run, and the results reveal that the best order for rule-based affix elimination is prefix, circumfix, suffix, and infix, whereas the Ahmad algorithm produces a better performance compared to the Othman algorithm.

Meanwhile, Sankupellay & Valliapan [24] developed the Mangalam algorithm, where they adopted the Porter stemming algorithm to stem Malay documents. Although the Porter stemmer is commonly used to stem English words, the algorithm is adaptable to handle dual words, or “kata ganda" from Malay documents. The Porter stemmer was successfully adopted in stemming Indonesian text [25]. The stemmer used a root word dictionary to validate the four categories of affixes, including inflection particles, possessive pronouns, derivation suffixes, and derivation prefixes. Some of the studies conducted by Rosid et al. [26] used the Sastrawi library in their studies to examine the comparison of stemming result against Tala porter stemmer. Tala porter stemmer was developed by Fadilah Z. Tala in 2003 using five steps in Porter stemmer by imitating how words are derived and inflected [27]. This study used 50 of the Indonesian student complaint documents. The findings of the study indicate that employing the Sastrawi dictionary yields superior results in comparison to utilising the Tala Porter dictionary. The result shows 92% accuracy when they used the Sastrawi dictionary, and 82% when they just used the Tala Porter stemmer. The processing speed for Sastrawi libraries is faster than Tala Porter with 0.6 seconds compared to 241.6 seconds for Tala Porter.

III. Research Methodology

The overall framework of research methodology in this study is depicted in Fig. 1. There are three main phases, including textual data collection, text preprocessing, and performance evaluation.

A. Phase 1: Textual Data Collection

This study makes use of textual datasets from the following two primary domains: education and housebreaking crime. The first textual dataset employed in this study consists of a compilation of course summaries in the Malay language. Meanwhile, the second dataset comprises a collection of housebreaking crime reports ranging from 2010 to 2013, also in the Malay language. The two datasets were obtained from a higher education institution and the Royal Malaysia Police Department, and they are both closed domain datasets. Table I shows the detailed descriptions of both textual datasets. For the purposes of this study, a sample of 100 course summaries and 100 housebreaking crime reports was randomly chosen and stored in Excel format. The collection of 100 documents of course summaries comprises a cumulative total of 5,520 words, whereas the set of 100 housebreaking crime reports contains a total of 3,530 words. The range of word lengths observed in the course summaries documents spans from 24 to 135 words, while the housebreaking crime reports consist of approximately 22 to 155 words.

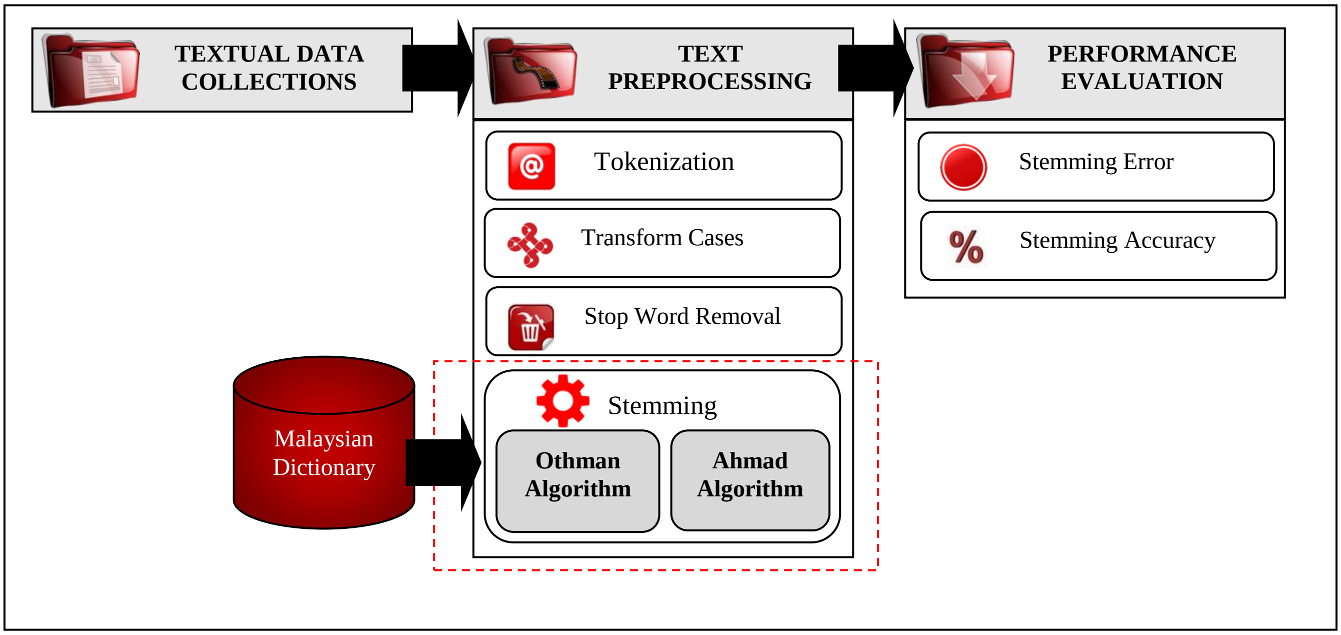


Fig. 1. Framework of research methodology.

TABLE I. DETAILS DESCRIPTION OF TEXTUAL DATASET FOR EDUCATION AND HOUSEBREAKING CRIME DOMAIN

Domain	Number of Documents	Total Number of Words	Range of Word Lengths
Education	100	5520	24-135
Housebreaking Crime	100	3530	22-155

B. Phase 2: Text Preprocessing

The text processing phase contains several steps, such as tokenization, transform cases, stop word removal, and stemming. These steps are essential to reducing the noise of the words in the raw dataset. It is crucial to reduce the document's feature size before proceeding to the next computational task. Tokenization processes are splitting the sentence or paragraph into single words, while transform cases are the process of converting all words to lowercase. A stop word refers to a word that is highly prevalent and frequently occurs in both written and spoken language but does not contribute significant semantic meaning to the overall document. Examples of stop words include prepositions, conjunctions, numbers, and punctuation marks. In Malay, examples of stop words are “di” (at), “ke” (to), “dengan” (with), and “dari” (from). In this experimental study, a total of 323 stop words were identified and subsequently employed to eliminate unnecessary words from the dataset. Table II presents a list of 100 stop words that were utilised in the context of this research.

TABLE II. AN EXAMPLE OF LIST OF STOP WORDS IN MALAY LANGUAGE

List of Stop Words				
ada	seandainya	kalau	apa-apa	sekitar
inikah	agar	sebelumnya	katakan	atau
sampai	janganlah	allah	segala	kepadaku
adakah	sebab	kami	apabila	selain
inilah	akan	sebenarnya	ke	ataukah
sana	jika	amat	sehingga	kepada
adakan	sebagai	kamikah	apakah	selalu
itu	aku	secara	kecuali	ataupun
sangat	jikalau	antara	sejak	kepadamu
adalah	keatas	kamipun	apapun	selama
itukah	akulah	sedang	kelak	bagaimana
sangatlah	jua	antaramu	sekalian	kepadanya
adanya	sebanyak	kamu	atas	diatas
itulah	akupun	sedangkan	kembali	di
saya	juapun	antaranya	sekalipun	samping
adapun	sebelum	kamukah	atasmu	seluruh
jadi	al	sedikit	kemudian	bagi
se	juga	apa	sekarang	kerana
agak	dari	kamupun	atasnya	seluruhnya
jangan	alangkah	sedikitpun	kepada	bagimu

The subsequent procedure involves the execution of the stemming process. Comparative experiments were conducted to evaluate the effectiveness of the Othman and Ahmad algorithms for stemming. Both algorithms are being implemented using the Java programming language. Fig. 2 depicts the Othman algorithm, while Fig. 3 depicts the Ahmad algorithm. The algorithms are combined with the proposed rules, with Othman employing 121 stemming rules (Set A) and Ahmad employing 432 (Set B) and 561 (Set C) stemming rules, respectively. Three distinct programmes have been developed to represent different sets of rules, in particular Othman Stemmer (Set A), Ahmad Stemmer (Set B), and Ahmad Stemmer (Set C). Meanwhile, six separate series of experiments were conducted to stem two different textual datasets from the education and housebreaking crime domains. The resulting stem words were stored in Excel files. The Malaysian dictionary used in this study consists of a collection of root words obtained from the Kamus Dewan Edisi Keempat.

```
Step-1: If there are no more words then stop, otherwise get the next word.
Step-2: If there are no more rules then accept the word as a root word and go
        to Step-1, otherwise get the next rule.
Step-3: Check the given pattern of the rule with the word: If the system finds
        a match, apply the rule to the word to get a stem word.
Step-4: Check the stem word against the dictionary; perform any necessary
        recoding and recheck the dictionary.
Step-5: If the stem word appears in the dictionary, then this stem word is the
        root of the word and go to Step-1. Otherwise go to Step-2.
```

Fig. 2. Othman algorithm [23].

```
Step-1: Find the next word till the last word.
Step-2: Check the word in the dictionary; if the word appears in dictionary, it
        is the root word and return to Step 1.
Step 3: Get the next rule; if no further rules are available, the word are root
        word and return to Step 1.
Step 4: Apply the rule to the word to get a stem word.
Step 5: Check the dictionary and recode for prefix spelling exceptions.
Step 6: If the stem word appears in the dictionary, it is root word and proceed
        to Step 1; otherwise go to Step 7.
Step 7: Examine the stem from Step 4 for spelling variations in the dictionary.
Step 8: If the word stem appears in the dictionary, it is root word and proceed
        to Step 1; otherwise go to Step 9.
Step 9: Check the dictionary and recode for suffix spelling exceptions.
Step 10: If the stem word appears in the dictionary, it is root word and
        proceed to Step 1; otherwise go to Step 3.
```

Fig. 3. Ahmad algorithm [23].

C. *Performance Evaluation*

After a series of experiments is run, the stemming result is collected for measuring the algorithm's performance. The performance value that has been considered in this study is the analysis of stemming errors, word stemming accuracy measurement, and processing speed. The analysis of stemming errors involves the examination of various types of errors, such as over-stemming, under-stemming, unstemmed words, and spelling exception errors. These errors have a negative impact on the efficacy of stemming algorithms. Over-stemming occurs when a larger portion of a word is cut off than is necessary, resulting in the incorrect reduction of an inappropriate stem word. As an example, the word “memakan” (eating) needs to be stemmed into “makan” (eat), but then the result gives only “mak” (mother) after stemming. Under-stemming, meanwhile, happens when the smaller portion of the word is stemmed into the inappropriate stem word. For instance, the word “pengadu” (informer) should be stemmed into “adu” (inform), but the result gives “gadu” (trigonostemon longifolius). For unstemmed words, there are no changes in the derivative word before stemming or after stemming. Spelling exceptions occur when a word is chopped off from the accurate affixes, resulting in the formation of different root words. For example, the word “pengawal” (guard) is stemmed into “gawal” (confuse) instead of “kawal” (control). The other category for stemming errors refers to results other than these four types.

Meanwhile, for measuring the stemming accuracy, The equation below depicts the formula. Correctly stemmed words are calculated by conducting a comparison with the manually stemmed data and the root word as listed in the Malaysian dictionary, Kamus Dewan Edisi Keempat.

$$Accuracy = \frac{Total\ of\ correctly\ stemmed\ word}{Total\ of\ correctly\ and\ incorrectly\ stemmed\ words}\, x\ 100\%$$

In regard to processing speed, the duration is measured subsequent to the completion of the stemming process. The experiment was conducted utilising the Java Programming language within the NetBeans Integrated Development Environment, Version 12.0. The stemmer programme is executed on a desktop computer equipped with an Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz 2.90 GHz processor and 16 GB of DDR3 memory.

IV. Result and Discussion

This section discusses the results of the experiments and the evaluation of the stemming process.

A. *Analysis of the Stemming Errors*

Despite the ongoing encouragement for the development of the Malay stemmer, there are still several challenges that need to be addressed. The analysis of stemming errors involves the examination of various types of errors, such as over-stemming, under-stemming, unstemmed words, and incorrect stemming.

Table III provides a summary of the total number of stop words that were eliminated from the education dataset and the housebreaking crime dataset. The implementation of the stop word removal process resulted in the elimination of 1,555 insignificant words in total from a corpus consisting of 100 course summaries. In the context of analysing 100 housebreaking crime reports, there are 864 frequently occurring words with no significant meaning that were eliminated. The stop word removal process results in a total word count of 3,965 for the course summaries dataset, while the housebreaking crime dataset retains 2,486 words.

TABLE III. A Summary of the Total Number of Eliminated Words After Stop Word Removal

Textual Dataset	Number of Removed Words	Total Number of Remaining Words
Course Summaries	1555	3965
Housebreaking Crime Reports	864	2486

Meanwhile, a comparative analysis of stemming errors generated by the Othman and Ahmad algorithms is presented in Table IV. Experiments are conducted in six series, with Set A containing 121 rules for the Othman algorithm, Set B containing 432 rules, and Set C containing 561 rules for the Ahmad algorithm. By using a collection of course summaries as the dataset in the education domain, the Othman algorithm correctly stems 1,582 words and identifies 1,643 root words. The Ahmad algorithm with Set B rules correctly stems 1,529 words, while the Ahmad algorithm with Set C rules correctly stems 1,526 words. In contrast to the Othman algorithm, which manages to identify only 1,643 root words, the Ahmad algorithm detects 1,959 root words for both sets. The Ahmad algorithm yields a higher number of under-stemmed words within the range of 168 to 174 in comparison to the Othman algorithm, which yields 130 under-stemmed words. In contrast, the Othman algorithm demonstrates a higher frequency of over-stemming errors, with 366 over-stemmed words, in comparison to the Ahmad algorithm, with 45 and 54 over-stemmed words for Set B and Set C, respectively. Meanwhile, the number of spelling exception errors generated by the Ahmad algorithm is greater in comparison to the Othman algorithm. Irrelevant words refer to root words that are not listed in the dictionary.

On the other hand, for the housebreaking crime dataset of the crime domain, the Othman and Ahmad algorithms for Set B and Set C achieve within the range of 736 to 737 words, where the difference between the two algorithms is not significantly different. However, with regard to root word identification, the Ahmad algorithm successfully detects 1,531 root words for Set B and Set C, while the Othman algorithm only detects 1,273 root words. In the context of under-stemming errors, the Ahmad algorithm yields a total of 113 under-stemmed words, which is higher than the count of 110 under-stemmed words produced by the Othman algorithm. In the meantime, the Othman algorithm also produced 285 over-stemmed words, which is higher than the Ahmad algorithm, which produced within the range of 21 to 22 over-stemmed words for both Set B and Set C. Meanwhile, the spelling exceptions generated by the Othman and Ahmad algorithms are within the range of 14 to 16 words.

TABLE IV. SUMMARY OF STEMMING ERRORS PRODUCED BY OTHMAN AND AHMAD ALGORITHMS FOR 100 COURSE SUMMARIES AND 100 HOUSEBREAKING CRIME REPORTS

Algorithms	**Othman Algorithm (SET A)**				**Ahmad Algorithm (Set B)**				**Ahmad Algorithm (Set C)**			
Dataset Domain	***Education***		***Crime***		***Education***		***Crime***		***Education***		***Crime***	
	#	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**
Correct Stemming	1582	39.90%	736	29.61%	1529	38.56%	736	29.61%	1526	38.49%	737	29.65%
No stemming, root word	1643	41.44%	1273	51.21%	1959	49.41%	1531	61.58%	1959	49.08%	1531	61.58%
Under-stemming	130	3.28%	110	4.42%	174	4.39%	113	4.55%	168	4.24%	113	4.55%
Over-stemming	366	9.23%	285	11.46%	45	1.13%	22	0.88%	54	1.36%	21	0.84%
Spelling exceptions	5	0.13%	14	0.56%	19	0.48%	16	0.64%	19	0.48%	16	0.64%
Irrelevant Words	239	6.03%	68	2.74%	239	6.03%	68	2.74%	239	6.03%	68	2.74%

When these two distinct dataset domains are compared, it is discovered that the Othman algorithm outperforms the Ahmad algorithm in correctly stemming 39.66% of the words in the education dataset. In contrast, the Ahmad algorithm has a higher success rate of 29.65% in correctly stemming words from the crime dataset compared to the Othman algorithm. The aforementioned contrast finding indicates that the efficacy of the stemming process is influenced not only by the variety of morphological rules, but also by the extent of vocabulary richness present in the datasets, which has the potential to impact the accuracy of stemming. Meanwhile, the Ahmad algorithm demonstrates superior performance in root word identification for the education and crime datasets, with respective efficiencies of 49.08% to 49.41% and 61.58%. This observation aligns with the technique employed by Ahmad's algorithm, wherein the search process is conducted on the dictionary prior to the application of the stemming rules.

In the context of under-stemming errors, the Ahmad algorithm for Set B produces higher under-stemmed words, which is around 4.39% for the education domain and 4.55% for the crime domain. These percentages represent the ratio of under-stemmed words to the overall count of words remaining subsequent to the elimination of stop words. This is the result of employing a dictionary lookup function, which verifies the existence of the root word in the stemmed word even after it has been stemmed by a single rule. In the situation where the stemmed word is present in the dictionary, the evaluation against subsequent stemming rules stops, and the word is deemed a root word. Meanwhile, the Othman algorithm produces a higher frequency of over-stemming errors, which is around 9.46% for the education dataset and 11.46% for the crime domain. This demonstrates that the diversity of morphological rules is crucial for reducing over-stemming errors in the Malay language, which has a complex morphological structure due to the presence of affixes.

B. Stemming Acuracy

The accuracy of stemming performance is assessed by computing the ratio of correctly stemmed words to the sum of correctly and incorrectly stemmed words. Correctly stemmed words are recognised as the word that has been truncated to the appropriate root word and the word that has been correctly identified as the root word. On the other hand, words that have been incorrectly stemmed are classified as either under-stemmed, over-stemmed, or words with spelling exceptions.

Table V and Fig. 4 present a comparative analysis of the performance accuracy achieved by the Othman and Ahmad algorithms in the domains of education and crime. The Ahmad algorithm with Set B rules, when applied to the education dataset consisting of 100 course summaries, demonstrates the highest level of accuracy in stemming performance, achieving a rate of 93.61%. This accuracy rate is only marginally different from the Ahmad algorithm with Set C rules, which achieves a rate of 93.53%. In contrast, the Othman algorithm yields the least accurate results, approximately 86.04%.

In the context of a crime dataset consisting of 100 housebreaking crime reports, the Ahmad algorithm, when employing Set C rules, achieves the highest level of accuracy in stemming performance, with a rate of 93.80%. However, the difference in stemming accuracy achieved by implementing Set B rules is minimal, with a mere 0.04% variation. Meanwhile, the Othman algorithm exhibits a comparatively lower level of accuracy, measuring at 83.09%.

TABLE V. COMPARATIVE ANALYSIS OF ACCURACY PERFORMANCE FOR OTHMAN ALGORITHM AND AHMAD ALGORITHM

Algorithms	**Education Domain**					**Crime Domain**				
	Correctly stemmed words		***Incorrectly stemmed words***	***Correctly stemmed words + Incorrectly stemmed words***	***Accuracy (%)***	***Correctly stemmed words***		***Incorrectly stemmed words***	***Correctly stemmed words + Incorrectly stemmed words***	***Accuracy (%)***
	Correct stemmed word	***Correct root word identified***				***Correct stemmed word***	***Correct root word identified***			
Othman Algorithm (SET A)	1572	1633	520	3725	86.04%	736	1273	409	2418	83.09%
Ahmad Algorithm (SET B)	1529	1959	238	3726	93.61%	736	1531	151	2418	93.76%
Ahmad Algorithm (SET C)	1526	1959	241	3726	93.53%	737	1531	150	2418	93.80%

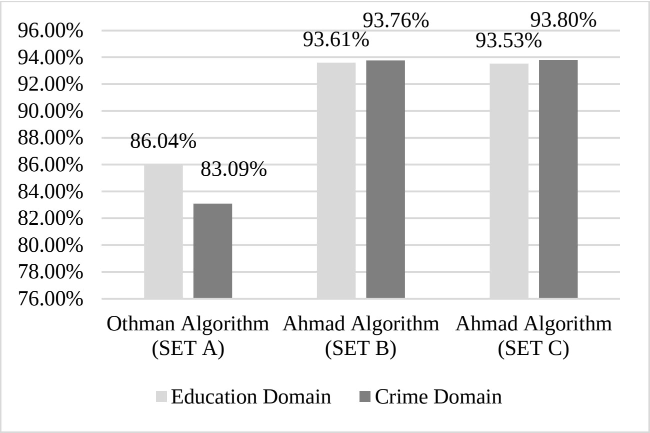


Fig. 4. Stemming accuracy for othman algorithm and ahmad algorithm.

The Ahmad algorithm consistently exhibits superior accuracy in stemming performance when compared to the Othman algorithm, irrespective of the nature of the datasets used. The findings show that the variants of morphological rules and the dictionary lookup approach are significantly contributing to the overall stemming accuracy.

V. CONCLUSION

The purpose of this study is to perform a comparative analysis of the stemming performance of the Ahmad and Othman algorithms, two prominent and pioneering Malay stemming algorithms, on housebreaking crime reports. The Ahmad algorithm has the greatest stemming accuracy rate, indicating that it is extremely unreliable for producing stem words across all dataset domains. The insignificant difference in stemming accuracy when Set B, which consists of 432 rules, and Set C that contains 561 rules, are implemented in the Ahmad algorithm indicates that the 432 rules are sufficiently significant to yield favourable outcomes, whether for stemming the dataset in the domain of education or crime. This is evident from the stemming accuracy results, which indicate a difference of approximately 0.08% and 0.04% between these two sets of rules applied to the education data and crime data set, respectively. Meanwhile, the dataset comprising textual records of housebreaking crime reports exhibits a substantial presence of root words, thereby requiring less effort for stemming operations. The performance of stemming accuracy is significantly impacted by this factor, as the quantity of root words present in the dataset directly correlates with the number of correctly stemmed words. There is a decreased probability of stemming errors occurring when fewer stemming operations are necessary. In addition, a restricted vocabulary range, and a notable prevalence of word repetition in the textual dataset also contribute to the stemming result. Henceforth, a thorough assessment of the functionalities of these two algorithms may be extended to encompass additional textual datasets that are more vocabulary-rich.

ACKNOWLEDGMENT

This research is supported by Ministry of Education Malaysia, under Fundamental Research Grant Scheme (FRGS) with vote number 59676 (Reference Code: FRGS/1/2021/ICT03/UMT/01/1).

REFERENCES

[1] Ittoo, L. M. Nguyen, and A. Van Den Bosch, “Text analytics in industry: Challenges, desiderata and trends,” Computers in Industry, vol. 78. Elsevier B.V., pp. 96–107, May 01, 2016. doi: 10.1016/j.compind.2015.12.001.

[2] S. J. Barnes, M. Diaz, and M. Arnaboldi, “Understanding panic buying during COVID-19: A text analytics approach,” Expert Syst. Appl., vol. 169, no. November 2020, p. 114360, 2021, doi: 10.1016/j.eswa.2020.114360.

[3] I. Feinerer, K. Hornik, and D. Meyer, “Text mining infrastructure in R,” J. Stat. Softw., vol. 25, no. 5, pp. 1–54, 2008, doi: 10.18637/jss.v025.i05.

[4] P. Carracedo, R. Puertas, and L. Marti, “Research lines on the impact of the COVID-19 pandemic on business. A text mining analysis,” J. Bus.

Res., vol. 132, no. September 2020, pp. 586–593, 2021, doi: 10.1016/j.jbusres.2020.11.043.

[5] S. Hassan, H. Mubarak, A. Abdelali, and K. Darwish, “ASAD: Arabic social media analytics and understanding,” EACL 2021 - 16th Conf. Eur. Chapter Assoc. Comput. Linguist. Proc. Syst. Demonstr., pp. 113–118, 2021, doi: 10.18653/v1/2021.eacl-demos.14.

[6] V. Y. Radygin, D. Y. Kupriyanov, R. A. Bessonov, M. N. Ivanov, and I. V. Osliakova, “Application of text mining technologies in Russian language for solving the problems of primary financial monitoring,” Procedia Comput. Sci., vol. 190, no. 2019, pp. 678–683, 2021, doi: 10.1016/j.procs.2021.06.078.

[7] S. V. Praveen, R. Ittamalla, and G. Deepak, “Analyzing the attitude of Indian citizens towards COVID-19 vaccine – A text analytics study,” Diabetes Metab. Syndr. Clin. Res. Rev., vol. 15, no. 2, pp. 595–599, 2021, doi: 10.1016/j.dsx.2021.02.031.

[8] N. Zhang, Q. Jia, K. Yin, L. Dong, F. Gao, and N. Hua, “Conceptualized Representation Learning for Chinese Biomedical Text Mining,” no. 1, pp. 1–4, 2020, [Online]. Available: http://arxiv.org/abs/2008.10813

[9] W. Chansanam and K. Tuamsuk, “Thai twitter sentiment analysis: Performance monitoring of politics in Thailand using text mining techniques,” Int. J. Innov. Creat. Chang., vol. 11, no. 12, pp. 436–452, 2020.

[10] A. Rizk and A. Elragal, “Data science: developing theoretical contributions in information systems via text analytics,” J. Big Data, vol. 7, no. 1, 2020, doi: 10.1186/s40537-019-0280-6.

[11] M. L. Carnot, J. Bernardino, N. Laranjeiro, and H. G. Oliveira, “Applying text analytics for studying research trends in dependability,” Entropy, vol. 22, no. 11, pp. 1–20, 2020, doi: 10.3390/e22111303.

[12] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, “Text preprocessing for text mining in organizational research: Review and recommendations,” Organ. Res. Methods, vol. 25, no. 1, pp. 114–146, 2022.

[13] R. A. Sinoara, J. Antunes, and S. O. Rezende, “Text mining and semantics: a systematic mapping study,” J. Brazilian Comput. Soc., vol. 23, no. 1, 2017, doi: 10.1186/s13173-017-0058-7.

[14] N. Wang, J. Zeng, M. Ye, and M. Chen, “Text mining and sustainable clusters from unstructured data in cloud computing,” Cluster Comput., vol. 21, no. 1, pp. 779–788, 2017, doi: 10.1007/s10586-017-0909-1.

[15] J. Singh and V. Gupta, A systematic review of text stemming techniques, vol. 48, no. 2. Springer Netherlands, 2017. doi: 10.1007/s10462-016-9498-2.

[16] M. N. Kassim, M. A. Maarof, A. Zainal, and A. A. Wahab, “Word stemming challenges in Malay texts: A literature review,” 2016 4th Int. Conf. Inf. Commun. Technol. ICoICT 2016, vol. 4, no. c, 2016, doi: 10.1109/ICoICT.2016.7571887.

[17] M. N. Kassim, S. H. M. Jali, M. A. Maarof, and A. Zainal, “Towards stemming error reduction for Malay texts,” Lect. Notes Electr. Eng., vol. 481, pp. 13–23, 2019, doi: 10.1007/978-981-13-2622-6_2.

[18] N. I. and S. M. F. D. S. MUSTAFA, “Stemming For Term Conflation In Malay Texts,” 2001.

[19] M. Yasukawa, H. T. Lim, and H. Yokoo, “Stemming malay text and its application in automatic text categorization,” IEICE Trans. Inf. Syst., vol. E92-D, no. 12, pp. 2351–2359, 2009, doi: 10.1587/transinf.E92.D.2351.

[20] M. M. N. M. Kassim, M. M. A. M. Maarof, A. Zainal, and A. A. Wahab, “Enhanced Affixation Word Stemmer with Stemming Error Reducer to Solve Affxation Stemming Errors,” J. Telecommun. Electron. Comput. Eng., vol. Vol 8, no. No. 3, pp. 37–41, 2016, [Online]. Available: http://journal.utem.edu.my/index.php/jtec/article/view/999

[21] M. Abdullah and F. Ahmad, “Rules frequency order stemmer for malay language,” … Int. J. …, vol. 9, no. 2, pp. 433–438, 2009, [Online]. Available: http://paper.ijcsns.org/07_book/200902/20090258.pdf

[22] R. Alfred, L. C. Leong, C. K. On, and P. Anthony, “A Literature Review and Discussion of Malay Rule - Based Affix Elimination Algorithms,” pp. 285–297, 2014, doi: 10.1007/978-94-007-7287-8.

[23] F. Ahmad, M. Yusoff, and T. M. T. Sembok, “Experiments with a stemming algorithm for Malay words,” J. Am. Soc. Inf. Sci., vol. 47, no. 12, pp. 909–918, 1996, doi: 10.1002/(SICI)1097-4571(199612)47:12<909::AID-ASI4>3.0.CO;2-6.

[24] M. Sankupellay and S. Valliappan, “Malay-language stemmer,” Sunw. Acad. J., vol. 3, pp. 147–153, 2006.

[25] M. A. Nazief, Bobby, “‘Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia.,’” Intern. Publ. Fac. Comput. Sci. Univ. Indones. Depok, Jakarta, 1996.

[26] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, “Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi,” IOP Conf. Ser. Mater. Sci. Eng., vol. 874, no. 1, 2020, doi: 10.1088/1757-899X/874/1/012017.

[27] F. Z. Tala, “A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia,” M.Sc. Thesis, Append. D, vol. pp, pp. 39–46, 2003.

Advanced Detection of COVID-19 Through X-ray Imaging using CovidFusionNet with Hybrid CNN Fusion and Multi-resolution Analysis

Majdi Khalid
Department of Computer Science and Artificial Intelligence-College of Computers, Umm Al-Qura University,
Makkah 21955, Saudi Arabia

***Abstract*—The rapid diagnosis of COVID-19 through imaging is crucial in the current pandemic scenario. This study introduces the CovidFusionNet, a novel model adapted for efficient COVID-19 image classification. By effectively combining fusing features from seven pre-trained convolutional neural networks (CNNs), our model presents better accuracy in detecting COVID-19 from X-ray images. Three separate datasets, obtained from Kaggle, were used in this study to ensure the reliability and robustness of the model. The Continuous and Discrete Wavelet Transform was implemented for robust multi-resolution image analysis to maintain image properties after denoising. A novel enhancement method was also proposed, combining the capabilities of Adaptive Histogram Equalization (AHE) and Wavelet Transforms to emphasize finer details and concurrently heighten clarity while minimizing noise. Furthermore, to mitigate class imbalance, an oversampling approach was implemented. Comprehensive validation using 12 metrics across each dataset verified the proposed consistent performance, with remarkable accuracies of 98.02% for Dataset One, 99.30% for Dataset Two, and 98.25% for Dataset Three. Comparing CovidFusionNet against seven well-known pre-trained models showed that CovidFusionNet appeared more capable. This research advances the area of image-based diagnosis using COVID-19 and provides a model for quick medical actions.**

Keywords*—*COVID-19 diagnosis; X-ray imaging; wavelet transform; Adaptive Histogram Equalization (AHE); oversampling; image denoising; image classification

I. Introduction

COVID-19 is a novel infectious disease caused by a new flu virus. It first emerged in Wuhan, China, in December 2019 [1]. It's one of the largest worldwide challenges of the 21st century [2, 3]. In March 2020, the World Health Organization (WHO) stated it was a widespread epidemic [4, 5]. COVID-19 is similar to other diseases like MERS and SARS since they all originate from the coronavirus family and harm our lungs [6, 7]. People with COVID-19 could cough, have a fever, feel tired, and lose their sense of taste and smell [8]. Some feel severely ill, have difficulties breathing, or even face life-threatening conditions, including kidney failure [9]. Many individuals died from it globally [10]. Thankfully, numerous companies have created vaccinations and various testing are being done over the world.

Finding persons with COVID-19 quickly is critically crucial to stop it from spreading. One approach to test for it is called RT-PCR [11, 12]. This test takes roughly four to six hours to display results. But, since so many people require testing, laboratories grow incredibly busy and cannot test everyone quickly enough [13]. This means a few people do not get diagnosed and may transmit the infection to others. We need a speedier, automated mechanism to test individuals. It would be beneficial to develop a simpler test that returns answers even sooner, so everyone can know whether they have the virus.

The rapid spread of the COVID-19 pandemic has necessitated the development of effective diagnostic tools for its timely detection. Chest X-ray imaging has emerged as a pivotal diagnostic method to assess the impact of the virus on the lungs. To enhance the efficiency and accuracy of detecting COVID-19 from X-ray images, researchers have proposed various computational methods, primarily harnessing the potential of deep learning. These models aim to swiftly identify patterns indicative of the virus, thereby aiding quicker clinical interventions. This literature review delves into multiple approaches undertaken by researchers worldwide, offering a comprehensive understanding of the advancement in this domain.

In research [14], Terry Gao and Grace Wang employed a set of lung X-ray images that were used to train a deep CNN that can distinguish between noise and useful information. This CNN can then use the data to train and interpret new images by spotting patterns that point to COVID-19. Singh et al. in study [15] suggested the Covid-aid model, an extension of the DarkCovidNet's architecture. With 19 convolution layers and six max-pooling layers, the model defines the lung X-ray images to determine normal, COVID-19 and pneumonia. Though the model could successfully classify the data, the classification accuracy is only 87%. Similarly, Shah et al. [16] proposed a hybrid model that employs a convolutional neural network (CNN) and Gated Recurrent Unit (GRU) to classify the diseases. The CNN model performs feature extraction in this study, whereas the GRU serves as an image classifier. The model is trained and validated with 200 epochs, and on the final epoch, the validation accuracy is as high as 93%. In [17], Khan et al. proposed a CNN architecture to classify COVID-19 pneumonia. The model is constructed employing the transform-merge block (STM) and RE-based operation for feature extraction. Among the three datasets used on the study, the proposed model performs best on the CoV-NonCoV-15k dataset with an accuracy rate of 96.53%.

To classify COVID-19 from chest X-ray images, Gayathri et al. in [18] have proposed an ensemble model after experimenting with several pre-trained model. Among the experimented models, the best outcome was observed from the fusion of the InceptionResNetV2 and Xception models with an accuracy of 95.78%. However, the outcome only depends on the conditions under which the Sparse autoencoder is used for dimensionality reduction and a Feed Forward Neural Network is employed for COVID-19 detection. Banerjee et al. [19] proposed a random forest meta-learning blending algorithm. The strategy follows the decision score technique. For feature extraction of the X-ray images, DenseNet201 architecture is implemented. However, the model only performs well when the dataset is small. On the small Chowdhury et al.'s dataset, the model showed an accuracy rate of 98.13%, whereas, on the larger Wang et al.'s COVID-X dataset, the accuracy was 94.55%.

In the paper [20], Ismael et al. has proposed multiple approaches to classify COVID-19 from the chest X-ray images using deep learning-based models. The study used several CNN models to extract features from the image data. With the assistance of ResNet50 deep feature extractor, the SVM classifier obtained a classification accuracy of 94.7%. Additionally, the authors suggested a fine-tuned CNN model that gives an accuracy rate of 92.6% for the same image data. However, in this study, the lowest accuracy of 91.6% was obtained from a CNN model with end-to-end training. It can be understood that deep approaches are more efficient in the classification of COVID-19 X-ray image data. In study [21], Kanjanasurat et al. used a combination of CNN and RNN techniques. In this study, the fully connected layers of the CNN model, ResNet152V2, were replaced by the RNN model, GRU, to achieve a classification accuracy of 93.37%. Here, the CNN layers were responsible for extracting the features of the data and calculation of dependencies and classification were performed by the RNN layers.

For the classification of chest X-ray images obtained from various sources, Alshmrani et al. suggested a VGG19 model with a fully connected network [22]. The model is assisted by CNN model for feature extraction. This technique provides an accuracy of 96.48% for image classification. In study [23] using ensemble method, Kuzinkovas et al. suggested a model where ANN, LR, LDA and RF performs the task of image classification with an accuracy of 98.34%. During the classification task, the ensemble model uses ResNet50, VGG19, VGG16, and GLCM for feature extraction. Another CNN-based model was suggested by Hafeez et al. in [24] for COVID-19 classification. The proposed CODSC-CNN is consisted of 8 weighted and two fully connected layers. The model is inspired by the pre-trained AlexNet and VGG16 models. The model has an 89% success rate in identifying COVID-19 X-ray images.

Based upon prior research (see Table I) in the area of COVID-19 image identification, we found a few drawbacks in the present approaches. Particularly, several of these models struggle with challenges relating to accuracy, the complexity of managing unbalanced image datasets, and the limitation of constrained data availability. In an attempt to solve these inadequacies and improve the diagnostic effectiveness, we developed the CovidFusionNet. Our proposed model combines the capabilities of multiple pre-trained models to offer higher performance and attain exact identification of COVID-19 in imaging data. By employing an ensemble method, CovidFusionNet seeks to create a new standard in terms of accuracy and durability in COVID-19 identification via imaging. The primary motivation of CovidFusionNet is to optimize the efficiency and precision of COVID-19 diagnosis by exploiting medical imaging, particularly X-ray images. This involves providing medical professionals with a dependable and effective method to detect COVID-19 cases. This is essential due to the fast virus transmission and the need for immediate action. The main contributions in this paper are as follows:

- Proposed a novel fusion model, CovidFusionNet that effectively utilizes the features of multiple pre-trained CNNs to improve COVID-19 image classification. This strategic fusion model assures CovidFusionNet heightened accuracy and flexibility in identifying COVID-19 cases.
- Introduced the use of the Continuous and Discrete Wavelet Transform for multi-resolution analysis. This transformation retains the image's pixel value distributions, ensuring features are preserved during denoising.
- Proposed a novel enhancement method that combines the strengths of AHE and Wavelet Transforms. The method focuses on intricate details, simultaneously enhancing and minimizing noise for optimal image clarity.
- Recognized the presence of class imbalances in the datasets and applied an over-sampling approach, ensuring unbiased model performance.
- We employed 12 evaluation metrics across three distinct datasets to assess the model's reliability and consistency. This thorough evaluation confirms our model's ability to consistently perform well across various datasets, underscoring its adaptability and reliability. Furthermore, we compared the performance of our proposed model with seven pre-trained models to demonstrate its superior capabilities and enhancements.

The structure of the remainder of the paper is organized as follows: Section II presents the data collection process, data preprocessing techniques, handling, and the overall methodology of the proposed framework. Section III delves into the findings and results obtained from applying the proposed method. Section IV provides a comprehensive discussion on these findings, exploring their implications and significance. Section V provides the paper's final section, summarizing the main results and presenting valuable perspectives on possible future research areas in this field.

II. Proposed Method

In this study, we utilized a systematic strategy to diagnose COVID-19 through imaging. Using three distinctive Kaggle-sourced X-ray datasets, our initial step was image denoising using Continuous and Discrete Wavelet Transforms.

Additionally, a unique improvement approach, merging Adaptive Histogram Equalization (AHE) and Wavelet Transforms, was applied for the enhancement of the image. Recognizing the difficulty of data imbalance, an oversampling approach was applied. The main component of our technique, the CovidFusionNet, fuses feature from seven pre-trained CNNs, achieving greater accuracy. Rigorous validation was undertaken across measures, and comparison analyses were made against renowned pre-trained models. The entire procedure is portrayed in Fig. 1.

TABLE I. A CONCISE OVERVIEW OF CONTEMPORARY DEEP LEARNING RESEARCH ON X-RAYS OF COVID-19

Reference	Dataset	Methodology	Limitations
2020 [14]	Middlemore Hospital data	CNN	Small dataset, training set is unknown, lack of details performance metrics
	Kaggle		
2021 [15]	Joseph Paul Cohen's GitHub repository	Covid-Aid	Limited and imbalanced dataset, accuracy can be improved.
	ChestX-ray8 database structured by Wang et al.		
2021 [16]	Joseph Paul Cohen's GitHub repository	CNN +GRU	Small and limited dataset, limit to perform on multiple data.
	Kaggle repository		
2021 [17]	CoV-Healthy-6k	STM-RENet	Inadequate data preparation and variability in image interpretation.
	CoV-NonCoV-10k		
	CoV-NonCoV-15k		
2022 [18]	Joseph Paul Cohen's GitHub repository	InceptionResNetV2 + Xception	Limited dataset and absence of comparative analysis of pre-trained methods.
	Paul Mooney's Kaggle repository		
2022 [19]	Wang et al.'s dataset (COVID-X)	Blended Ensemble	Accuracy can be enhanced, insufficient preprocessing, and lack of model assessment.
	Chowdhury et al.'s dataset		
2021 [20]	Github, 2020	Finetuned ResNet50, CNN, ResNet50 + SVM	Model performance can be enhanced, and imbalance dataset.
	Kaggle, 2020		
	Radiology Assistant 2020		
2023 [21]	Joseph Paul Cohen's GitHub repository	ResNet152V2 + GRU	Imbalance dataset, lack of existing model evaluation, and comparatively low accuracy rate.
	Chowdhury et al.'s dataset Kang's dataset		
	Kermany's dataset		
2023 [22]	Various Public datasets	VGG19 + CNN	Inadequate model evaluation
	RSNA + SIRM + Radiopaedia		
	Various research articles		
2023 [23]	COVID-QU-Ex dataset	ANN+ LR+ LDA+RF	The absence of diverse datasets for model assessments
2023 [24]	COVID19 dataset (2020)	CNN	Decreased accuracy with varied datasets
	Kaggle repository		

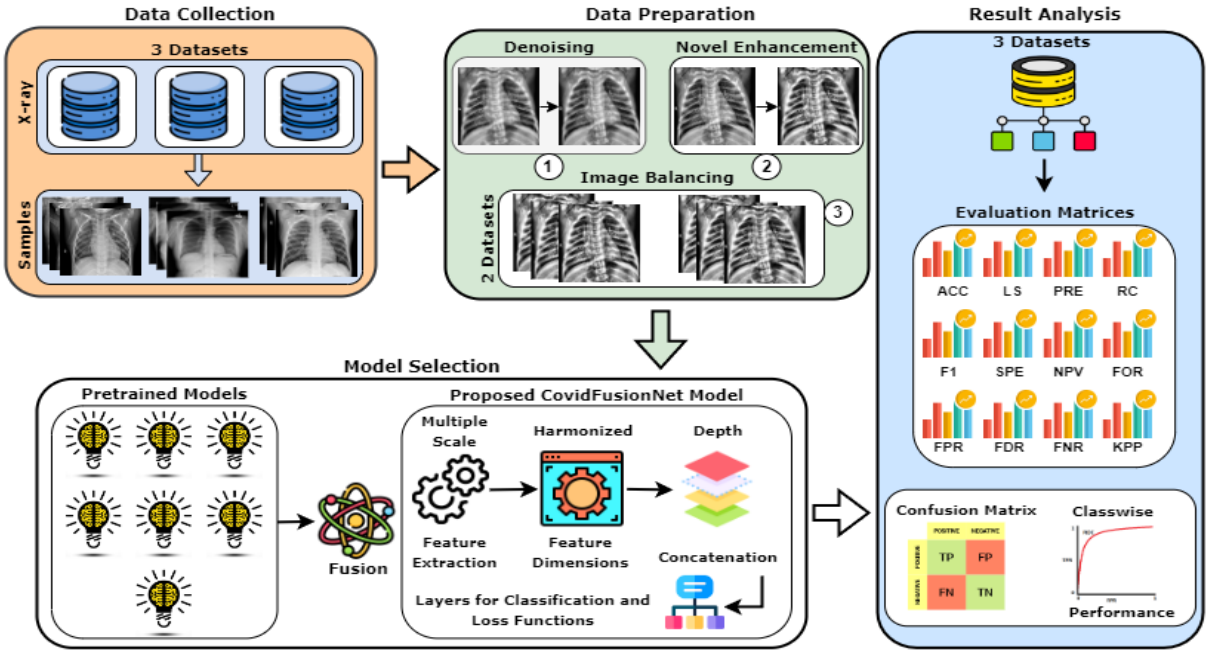


Fig. 1. Proposed workflow of our study for COVID-19 detection.

A. Data Dimensions

For the study on COVID-19 identification, we acquired three different datasets from Kaggle. The first dataset comprises 4,551 X-ray images. Among them, 1,281 images are COVID-19, whereas the remaining 3,270 are normal. This dataset is a compilation of COVID-19 Chest X-ray images gathered by aggregating 15 publicly accessible datasets. The second dataset gives us 4,626 images, split equally, with 2,313 indicating COVID-19 abnormalities and the other 2,313 being normal. The X-ray data used in this second dataset were obtained from numerous sources, such as the GitHub repository, Radiopaedia, the Italian Society of Radiology (SIRM), and the Figshare data repository sites. Our third and final dataset contains 2,159 X-ray images, out of which 576 show COVID-19 features, while 1,583 are categorized as normal. Fig. 2 displays the samples of three datasets. To improve our study's accuracy, we implemented several preprocessing methods. These comprised Image Denoising, Image Enhancement, and Image Balancing. These processes helped make the images more apparent, ensure they were accurately recognized, and make the data more consistent.

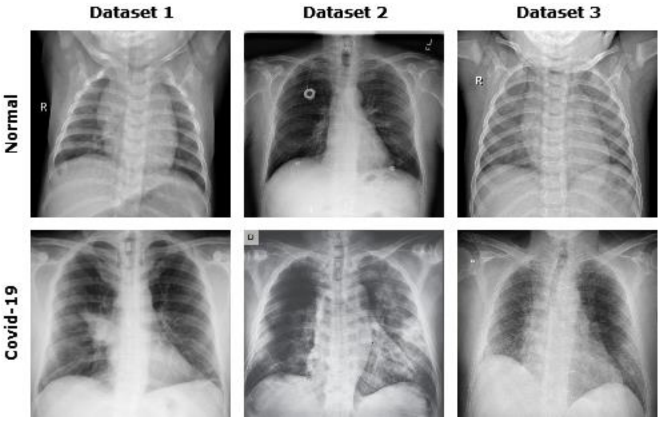


Fig. 2. Sample images from all three datasets.

B. Data Preparation

Data preparation is crucial since it eliminates inconsistencies and inaccuracies, assuring data accuracy. It prepares data for study by reducing noise and irrelevant information [25]. Preprocessing assists in discovering patterns that could be hidden in raw data. Ultimately, it boosts the performance and reliability of deep learning models.

1) Wavelet Transformation (Image Denoising): The Wavelet Transform enables a multi-resolution analysis of a given function or signal by describing it in terms of basic functions obtained from the dilation and translation of a prototype function, nicknamed the "mother wavelet," indicated a $\Psi(t)$. For the Continuous Wavelet Transform (CWT), the transform of a function $f(t)$ can be described as:

$$\boldsymbol{W_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\Psi(\frac{t-b}{a})dt} \qquad (1)$$

Here, $\mathrm{W_f(a,b)}$ denotes the wavelet coefficient, with a being the dilation parameter and b the translation parameter. The transformation process encompasses the complete domain of $\mathrm{f(t)}$. Crucially, the selected mother wavelet $\Psi(\mathrm{t})$ must conform to the admissibility requirement, defined as:

$$\boldsymbol{C_\Psi = \int_{-\infty}^{\infty}(\frac{|\Psi^\wedge(f)|^2}{|f|})df < \infty} \qquad (2)$$

In this instance, $\Psi^\wedge(f)$ stands for the Fourier transform of $\Psi(t)$. When transferring to the domain of discrete signals, especially crucial for digital applications, the Discrete Wavelet Transform (DWT) becomes increasingly significant. In the DWT, the continuous parameters a and b are discretized, generally specified as $a = 2^m$ and $b = n2^m$, with m and n being integers. This discretization allows a hierarchical examination of the function. The DWT is applied in both row and column directions for two-dimensional data such as images. This leads to four sets of coefficients capturing different information: the approximation (LL) , horizontal (LH), vertical (HL), and diagonal details (HH). A fundamental virtue of the wavelet transform is its reversibility, allowing for the original function or image to be rebuilt from its wavelet coefficients using the inverse wavelet transform. Fig. 3 illustrates the output of the Wavelet Transformation with its histogram. The histograms of the original and reconstructed images are remarkably similar, it shows that the wavelet transform (and its inverse) has successfully retained the pixel value distributions of the original image. This positive indicator shows that the wavelet transformation has kept the image's features well.

Algorithm 1. Hybrid AHE-Wavelet Image Enhancement (HAWIE)

```
Procedure HAWIE (Image $I$, SavePath S)
  Load necessary libraries: numpy, OpenCv,
  PyWavelets, Matplotlib
  function AHE_Enhancement (Image)
      Apply Adaptive Histogram Equalization on the Image
    return Enhanced image using AHE
  end function
  function Wavelet_Enhancement (Image)
    Convert image to floating-point representation
    Decompose image into wavelet coefficients
    Modify wavelet coefficients
    Reconstruct image from modified coefficients
    return wavelet-enhanced image
  end function
  $I_{AHE}$ = AHE_Enhancement (Image $I$)
  $I_{Enhanced}$ = Wavelet_Enhancement ($I_{AHE}$)
  if SavePath S is provided then
    Save $I_{Enhanced}$ to S
  end if
  return $I$, $I_{Enhanced}$
end procedure
```

2) Hybrid AHE-Wavelet Image Enhancement (HAWIE): In this study, we proposed a Hybrid AHE-Wavelet image enhancement (HAWIE) method as a novel approach for image improvement. This approach combines the comprehensive contrast refinement of Adaptive Histogram Equalization (AHE) with the multi-resolution capabilities of Wavelet

Transforms. Specifically, AHE concentrates on very small overlapping regions of an image, ensuring that every microscopic feature is properly accentuated, whether in a bright or dark location. Following the AHE process, the Wavelet Transform takes center stage, breaking the image into an array of frequency components. This enables a more focused enhancement, where certain visual elements may be enhanced while any undesired noise is concurrently minimized. Our findings, as illustrated in Fig. 4, give an impressive visual illustration of this hybrid method. The difference image, in particular, emerges as a useful tool, clearly emphasizing places that have received adjustments or upgrades. This visualization becomes even more informative when matched with the accompanying scatter plot, where the X-axis indicates pixel values from the original image. At the same time, the Y-axis exhibits those from the improved image. Each point on this figure encapsulates a pixel, its position demonstrating the association between the original and enhanced pixel values. The deviations from the "Line of Identity", a red dashed line showing identical pixel values in both images, give essential insights into the transformational power of the HAWIE approach, emphasizing its capabilities, particularly in vital sectors such as medical imaging. Algorithm 1 provides the overview structure of HAWIE.

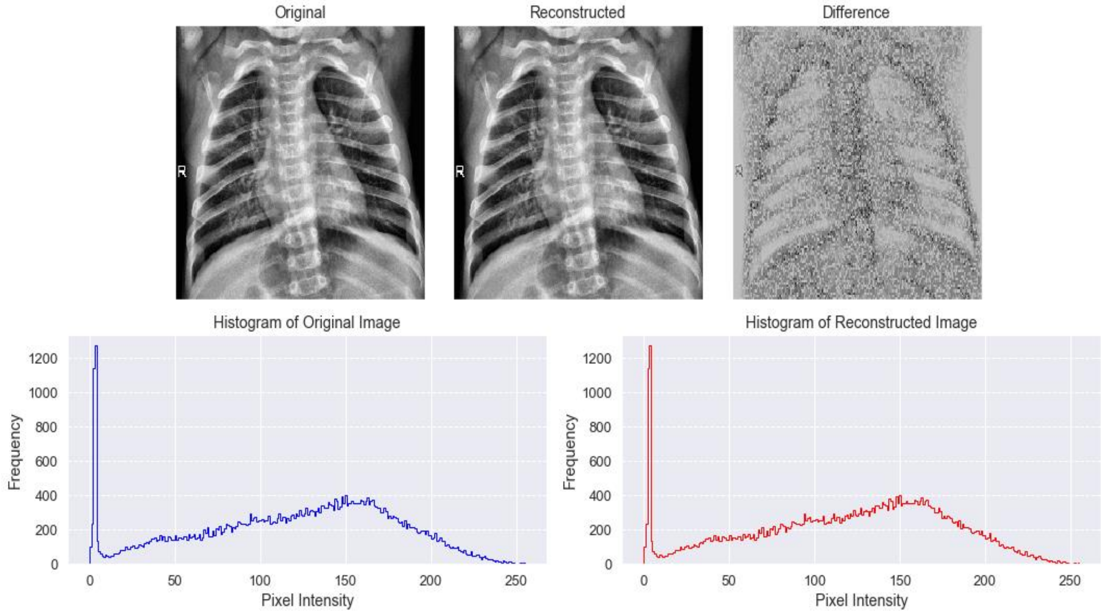


Fig. 3. Output of the wavelet transform with the histograms of the original and reconstructed images. The histogram comparison highlights the wavelet transform's efficacy in maintaining the image's features.

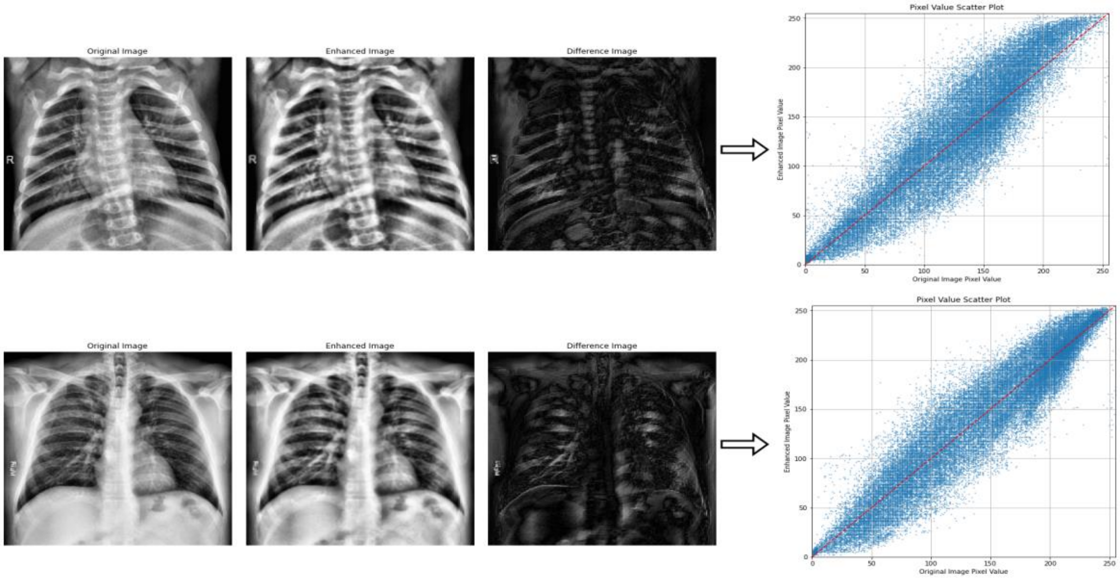


Fig. 4. Illustration of the HAWIE technique effects. The difference image reveals regions of enhancement, while the scatter plot showcases the relationship between original and enhanced pixel values, with deviations indicating modifications.

3) Image Balancing (Oversampling): In this study, we utilized three datasets for analysis. While Dataset 2 was already balanced, Datasets 1 and 3 demonstrated class imbalance, where some classes contained fewer samples than others. This mismatch may lead to biased model performance, perhaps favoring the more dominant classes. We utilized an oversampling approach to solve this difficulty, particularly for the minority classes in Datasets 1 and 3.

The oversampling approach operates by first selecting the class with the maximum number of samples, known as max_count. This count then sets the aim for all classes in the balanced dataset. For those classes with samples less than max_count, the oversampling strategy duplicates their samples until this goal number is attained. Replication is cyclic, meaning that once the end of a class's sample list is reached, it returns to the beginning, assuring a smooth continuation until the target sample count is attained. By applying this strategy, every class in the enlarged datasets (Datasets 1 and 3) now has an equal representation. This parity creates a balanced basis for subsequent data modeling and analysis. By executing this oversampling programmatically, the balanced datasets were stored in a different directory, guaranteeing clarity for later operations. The ultimate purpose of this technique is to reinforce the models' generalizability, trained on the modified datasets, ensuring they give unbiased insights unaffected by the over-representative classes of the original datasets. Table II represents the final datasets after all preprocessed attempts.

TABLE II. DISTRIBUTION OF SAMPLES ACROSS CLASSES IN DATASETS 1, 2, AND 3, ILLUSTRATING POST-OVERSAMPLING

Parameters	Dataset 1	Dataset 2	Dataset 3
Wavelet Transformation	✓	✓	✓
HAWIE	✓	✓	✓
Oversampling	✓	--	✓
Total Images	6540	4626	3166
Train (70%)	4578	3241	2216
Test (20%)	1308	925	633
Validation (10%)	654	460	317

C. Model Selection

For our study, we methodically selected seven pretrained models, each known for particular abilities in image processing. We included VGG19 for its comprehensive feature extraction from its deep 19-layer architecture. MobileNetV2 was selected for its best balance between computing performance and accuracy, which is appropriate for real-time activities. AlexNet, a pioneering model from the ImageNet competition, offers a basic baseline. Simultaneously, the innovative residual blocks of ResNet50, the multi-scale feature capture of InceptionV3, the fine-grained classifications from DenseNet201, and the efficiency of Xception each bring unique value. Together, this collection provides for complete performance evaluations across diverse architectures. Building on their distinct strengths, we next combined the special features of all seven deep learning models to develop our proposed innovative fusion model: CovidFusionNet.

1) VGG19: VGG19 is a well-known deep-learning architecture for classification of images [26]. It has 19 layers in total, including 16 convolutional layers, three fully connected levels, and five max-pooling layers. The persistent use of tiny 3×3 convolutional filters with a stride of 1, which are effective at extracting detailed information when stacked in succession, is a distinguishing characteristic of VGG19. VGG19 ensures the capture of detailed patterns at different spatial hierarchies by gradually increasing the number of filters from 64 in the initial layers to 512 in the deeper ones. The convolutional layers are separated by max-pooling layers with a 2×2 filter size and a stride of 2, which down-samples the spatial dimensions while retaining essential information. Following the convolutional layers, there are three fully connected layers, each having 4096, 4096, and 1000 neurons. Each convolutional layer has a ReLU activation to include nonlinearity in the model. While the uniform structure minimizes architectural complexity, the depth of VGG19 necessitates large processing resources. Nonetheless, its architecture has significantly impacted the field, emphasizing the promise of deeper networks for improved picture identification.

2) MobileNetV2: MobileNetV2 is a simplified deep-learning architecture designed for mobile devices and computationally constrained applications [27]. Inverted residuals and linear bottlenecks are central to its architecture, where channels are first reduced using a 1×1 convolution, followed by a 3×3 depthwise convolution, and then enlarged again, optimizing computational efficiency without compromising feature extraction capabilities. The model makes use of depthwise separable convolutions to reduce parameter counts by separating spatial and channel-wise calculations. MobileNetV2 slims its design even further by preceding fully connected layers in favor of average pooling leading into a SoftMax layer, giving it the highest level of efficiency and performance in on device deep learning applications.

3) AlexNet: AlexNet, a significant deep learning architecture, was essential in advancing the deep learning domain, particularly in the classification of images [28]. The network is composed of five convolutional layers, which are followed by three fully connected layers. The use of bigger filter sizes in AlexNet's first layers is a distinguishing feature; particularly, the first convolutional layer has 11×11 kernels with a stride of 4. As the network grows, it utilizes smaller and smaller filters, such as 5×5 and 3×3. AlexNet was among the first to use ReLU as activation functions, overcoming the vanishing gradient issue that afflicted networks that used classic sigmoid activations. Another novel idea introduced was the notion of dropout layers, in which a part of a neuron is randomly deactivated during training to prevent overfitting. The network also used data augmentation methods such as image translations and horizontal flips to increase the quantity and variety of the training dataset, which improved model generalization. AlexNet's exceptional performance entrenched deep neural networks as the dominant technique for picture classification, setting the framework for future architectural improvements in deep learning.

4) ResNet50: ResNet50, a member of the Residual Network group, transformed deep learning architectures by introducing residual connections, which are referred to as "skip connections." These connections enable the output of one layer to skip one or more following layers before being summed with the output of the latter, so minimizing the vanishing gradient issue and allowing the training of even deeper networks [29]. ResNet50 is made up of 50 layers that are organized as a blend of convolutional and identity blocks. Each block generally has three convolutional layers with filter sizes of 1×1, 3×3, and 1×1, which are used to decrease dimensionality, collect spatial data, and then restore dimensionality. Strides vary depending on the location of the layer, with stride-2 convolutions used to minimize spatial dimensions, half the height and length while doubling the number of filters. One of ResNet50's key characteristics is its capacity to retain information from previous layers, allowing it to capture both low-level and high-level features. Throughout the network, batch normalization and ReLU activation are employed to stabilize training and induce non-linearity. ResNet50's novel architecture, designed for high accuracy in image classification tasks, not only aids deeper model training but also establishes a new benchmark in deep learning, driving further design advances.

5) InceptionV3: InceptionV3 is a well-known deep learning model for image categorization. Its advanced "modules" execute concurrent convolutional processes with various kernel sizes, such as 1×1, 3×3, and 5×5, inside a single layer [30]. This multi-path technique allows InceptionV3 to record various spatial feature hierarchies at the same time, incorporating both granular and larger viewpoints. The use of factorized convolutions, which split bigger filters into smaller, asymmetric ones like 1×3 and 3×1, is one of InceptionV3's innovations, assuring computational efficiency without losing receptive scope. "Auxiliary classifiers" are deliberately put in the network's intermediate layers to push gradients to inner layers and boost regularization. Batch normalization is used consistently throughout the layers, giving stability to the activations and allowing for smoother training dynamics. Despite its sophisticated design, which combines multi-scale feature extraction with computational prudence, InceptionV3 exemplifies what is possible in terms of balancing accuracy and resource needs in the area of deep image classification.

6) DenseNet201: DenseNet201 distinguishes itself among deep learning architectures developed for image categorization by virtue of its remarkable dense connectivity [31]. Unlike the traditional technique, in which each layer gets information only from its immediate predecessor, DenseNet201 guarantees that each layer receives input from all previous levels, resulting in an extensive network of connections. This configuration encourages feature reuse and ensures more efficient gradient flow across the network, solving issues such as the disappearing gradient problem, which is common in deep architectures. Each layer, which is divided into interconnected blocks, generally employs 3×3 convolutional filters, with precise growth rates defining the insertion of additional filters as the network deepens. To regulate and aggregate feature-map dimensions, transition layers of 1×1 convolutional filters with a stride of 2 are alternated between these dense blocks. DenseNet201, with its 201 layers, is capable of collecting sophisticated image patterns without a massive rise in parameters due to its dense design. DenseNet201's architecture, which emphasizes continuous feature propagation and efficient gradient distribution, makes it highly resistant to overfitting and places it as a model of choice for complex image classification tasks.

7) Xception: Xception, which stands for "Extreme Inception," is a complex classification architecture that redefines the traditional inception technique by using depthwise separable convolutions [32]. This fundamental breakthrough divides convolutional processes into spatial and channel-based tasks, maximizing efficiency. The Xception model is composed of 71 layers that are organized into modules. Unlike standard Inception modules, which utilize a variety of filter sizes, Xception's architecture relies heavily on 3x3 convolutional kernels for depthwise operations, ensuring a thorough extraction of spatial data. 1x1 convolutions are utilized for cross-channel operations inside these modules. The model achieves strategic downsampling by adopting strides of 2 in selected modules. Xception incorporates residual connections, similar to the ResNet structure, to maintain smooth gradient flow throughout its depth, further strengthening its resilience. Xception proposes an architecture that delivers outstanding performance on picture classification tasks while assuring computational economy by combining the ideas of depthwise separable convolutions, intentional kernel choices, and a thoughtfully built layering scheme.

8) CovidFusionNet (Proposed Model): In this study, we proposed a novel CovidFusionNet model to increase classification performance, particularly for differentiating images related to COVID, by combining the strength of numerous pre-trained convolutional neural networks (CNNs) and their feature extraction capabilities.

The CovidFusionNet model begins by representing each image as I with a shape of 224×224×3. For each base model m within our ensemble model M (which encompasses seven models), the image undergoes a transformation through function f_m, generating a feature map F_m according to the equation 3:

$$F_m = f_m(I) \tag{3}$$

A harmonizing step is essential due to the varied spatial dimensions of the feature maps yielded by the different architectures. A resizing function, represented as R, refines each feature map to achieve a uniform shape:

$$F'_m = R(F_m) \tag{4}$$

Resultantly, F'_m is consistent with a dimension of 224×224×d_m, where d_m marks the depth as per model m. The essence of CovidFusionNet lies in its strategic feature fusion. Feature maps from all models are concatenated in depth, forming an integrated tensor F_c:

$$F_c = Concatenate\ (F'_{vgg16}, F'_{mobilenetv2}, \dots\dots, F'_{xception}) \quad (5)$$

After this fusion, F_c undergoes flattening and is channeled through a series of dense layers. Initially, it interacts with a weight matrix W_1 and bias b_1 to produce an output:

$$o_1 = ReLU\ (v.W_1 + b_1) \quad (6)$$

Post the application dropout regularization for model robustness, the final output is calculated as:

$$O = Softmax\ (o_1.W_2 + b_2) \quad (7)$$

The Categorical Cross entropy loss L supervises the training of the CovidFusionNet:

$$L = -\sum_{i=1}^{2} Y_i\ log\ (O_i) \quad (8)$$

This ensures optimal weight adjustments during training, directing the model toward accurate COVID-19 detection. Algorithm 2 depicts the operation of the proposed model.

Algorithm 2. Proposed CovidFusionNet Development and Operation

```
**Procedure** CovidFusionNet(Image $I$)
    Load pre-trained models into set $M$
    Freeze the weights of the models in $M$
    **for** each model $m$ in $M$ **do**
        Transform image $I$ using $m$ to get feature map $F_m$
        **if** $F_m$ does not have dimension 7×7×$d_m$
            Resize $F_m$ to $F'_m$ of dimension 7×7×$d_m$
        **else**
            $F'_m = F_m$
        **end if**
        Store $F'_m$ in list of feature maps $F'$
    **end for**
    Concatenate feature maps in $F'$ along depth to get $F_c$
    Flatten $F_c$ to get 1D tensor $T$
    $O_1$ = Dense layer with ReLU activation on $T$
    Apply dropout on $o_1$
    $O$ = Dense layer with Softmax activation on result
    **return** $O$
**end procedure**
```

The CovidFusionNet model exhibits remarkable improvements over to prior methodologies in medical image classification, specifically for identifying COVID-19. The main benefit of this approach is the use of numerous pre-trained convolutional neural networks (CNNs), which allows for a more extensive and diverse extraction of features compared to models that rely on a single CNN architecture. The variety of methods used to extract features leads to a classification that is less vulnerable to errors and more precise, which is essential for the specific needs of medical imaging. CovidFusionNet specifically tackles the issue of spatial dimension variability, which is a common occurrence when merging information from various CNNs. The resizing function is utilized to normalize feature maps prior to fusion, promising a consistent and harmonic integration of features. This strategy addresses the constraints associated with the restricted feature representation commonly observed in single-model approaches.

III. Performance Evaluation

Several evaluation metrics were examined to evaluate the efficacy of various deep-learning models. These measures, particularly Accuracy, Loss, Precision, Recall, F1-Score, Specificity, NPV, FOR, FPR, FDR, FNR, and Kappa Score, highlight the efficiency and dependability of the models under assessment. Fig. 5 to Fig. 10 presents a visual illustration of these scores. These figures show an in-depth visual representation of performance metrics [33] for the deep learning models analyzed in our study, including proposed CovidFusionNet (CvNet), VGG19 (VGG), MobileNetV2 (MV2), AlexNet (AxNT), RestNet50 (RsNT), Incep-tionV3 (InV3), DenseNet201 (DSNT), and Xception (XPTN). The metrics are placed across two separate rows inside the image, covering a wide variety of assessment criteria.

A. Performance Evaluation

To evaluate the efficacy of our employed classifiers in identifying COVID-19 from X-ray images, we applied a complete set of evaluation metrics (9) to (20). Where TP, FP, TN, and FN stand for the number of true positive, false positive, true negative and false negative cases, respectively.

Accuracy: Proportion of all X-ray images correctly identified, be it COVID-19 positive or negative.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (9)$$

Loss: Quantifies how well the prediction model performs compared to the actual outcomes.

$$Loss = -(y\ log(p) + (1-y)\ log(1-p)) \quad (10)$$

Precision: The percentage of X-rays that were accurately diagnosed as COVID-19 positive.

$$Precision = \frac{TP}{(TP+FP)} \quad (11)$$

Recall: The proportion of true COVID-19 positive X-rays that the model identified.

$$Recall = \frac{TP}{(TP+FN)} \quad (12)$$

F1-Score: Harmonic mean of precision and recall, ensuring a balance between them.

$$F1_{score} = 2 \times \frac{Precision*Recall}{Precision+Recall} \quad (13)$$

Specificity: The percentage of COVID-19 negative X-rays that the model accurately detected out of all the actual images.

$$Specificity = \frac{TN}{(TN+FP)} \quad (14)$$

NPV (Negative Predictive Value): The percentage of X-rays that were appropriately diagnosed as COVID-19 negative.

$$NPV = \frac{TN}{(TN+FN)} \quad (15)$$

FOR (False Omission Rate): percentage of COVID-19 positive cases that were actual but were misclassified as negative by the model.

$$FOR = \frac{FN}{(FN+TN)} \quad (16)$$

FPR (False Positive Rate): The proportion of real COVID-19 negative cases that were misclassified as positive by the model.

$$FPR = \frac{FP}{(FP+TN)} \quad (17)$$

FDR (False Discovery Rate): The proportion of X-rays mistakenly classified as COVID-19 positive but truly negative.

$$FDR = \frac{FP}{(FP+TP)} \quad (18)$$

FNR (False Negative Rate): The percentage of real COVID-19 positive cases that the model failed to predict.

$$FNR = \frac{FN}{(FN+TP)} \quad (19)$$

Kappa Score: Measures validity of predicted and actual classifications, adjusting for chance. A score near to 1 indicates outstanding alignment.

$$K = \frac{P_o - P_e}{1 - P_e} \quad (20)$$

where, p_o is the observed agreement, and p_e is the expected agreement.

B. Performance for Dataset 1

From the analysis of dataset 1, CvNet appears as a highly effective classifier, obtaining an ACC of 98.02%. Specifically, with CvNet, we notice an excellent PRE of 98.5%. It is followed at some distance by MV2, which records a score of 91.5%. Moving to the REC criteria, RsNT and InV3 exhibit impressive scores of 86.5%, but CvNet stands out with a significant score of 97.5%. Concerning F1S, whereas XPTN remains with a score of 82.75%, CvNet increases by reaching its highest point with a score of 98%. In terms of Specificity, a parameter important for understanding the real negative rate, CvNet again outperforms with 98.5%, with most other models.

The NPV, FOR, FPR, FDR, FNR, and the Kappa Score further show the capabilities of each model. Notably, CvNet differentiates out across these criteria. With an NPV of 98, it greatly beats VGG, which gets 84.5%. In analyzing the FOR, both VGG and InV3 score on the upper side at 15.5% and 15%, respectively. In comparison, CvNet has a respectable low score of 2%. Focusing on FPR, models like VGG, RsNT, and XPTN indicate heightened values of 15%, 13%, and 17%. Yet again, CvNet excels with a minimum 1.5%. As for FDR, XPTN's rate of 16.5% is in significant contrast to CvNet's low 1.5%. On the FNR front, although XPTN achieves a high 18%, CvNet displays its efficiency with only 2.5%. Lastly, considering the Kappa Score, CvNet's quality shines with 0.96, implying an excellent match with actual labels. On the other hand, XPTN scores 0.66, with MV2 and DSNT coming in between 0.82 and 0.78, respectively. Fig. 5 shows the performance of the models on dataset 1.

Fig. 6 shows the confusion matrix for multiple models, including CvNet, VGG, MV2, AxNT, RsNT, InV3, DSNT, and XPTN, when discriminating between COVID-19 and normal situations. For instance, evaluating the CvNet model, it can be determined that out of the COVID-19 instances, 638 were properly categorized as COVID-19 (True Positives) and 16 were wrongly classified as normal (False Negatives). On the other hand, when evaluating the normal cases, 644 were properly recognized as normal (True Negatives), and 10 were misclassified as COVID-19 (False Positives). Similar classification accuracy and error patterns were also found for the other models. For instance, the VGG model accurately recognized 549 COVID-19 samples but misclassified 105 as normal. Simultaneously, 556 normal samples were identified correctly, whereas 98 were misdiagnosed as COVID-19. However, the performance scores indicate that the CvNet model surpassed all other models in the classification performance for dataset 1.

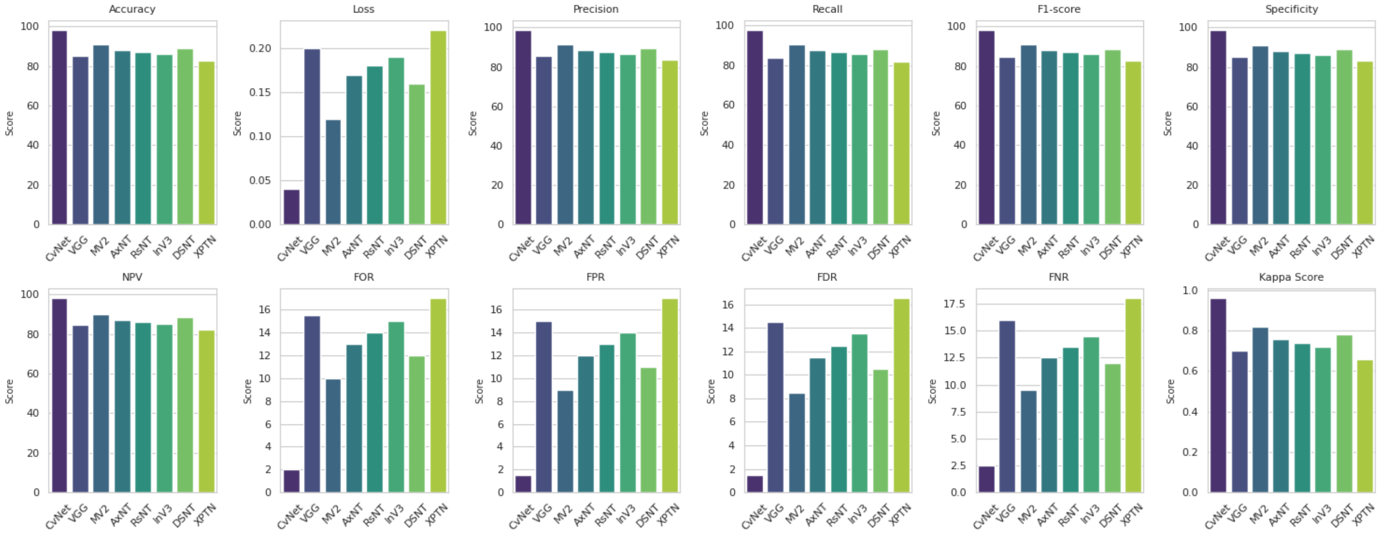


Fig. 5. Comparative performance metrics across the models on dataset 1.

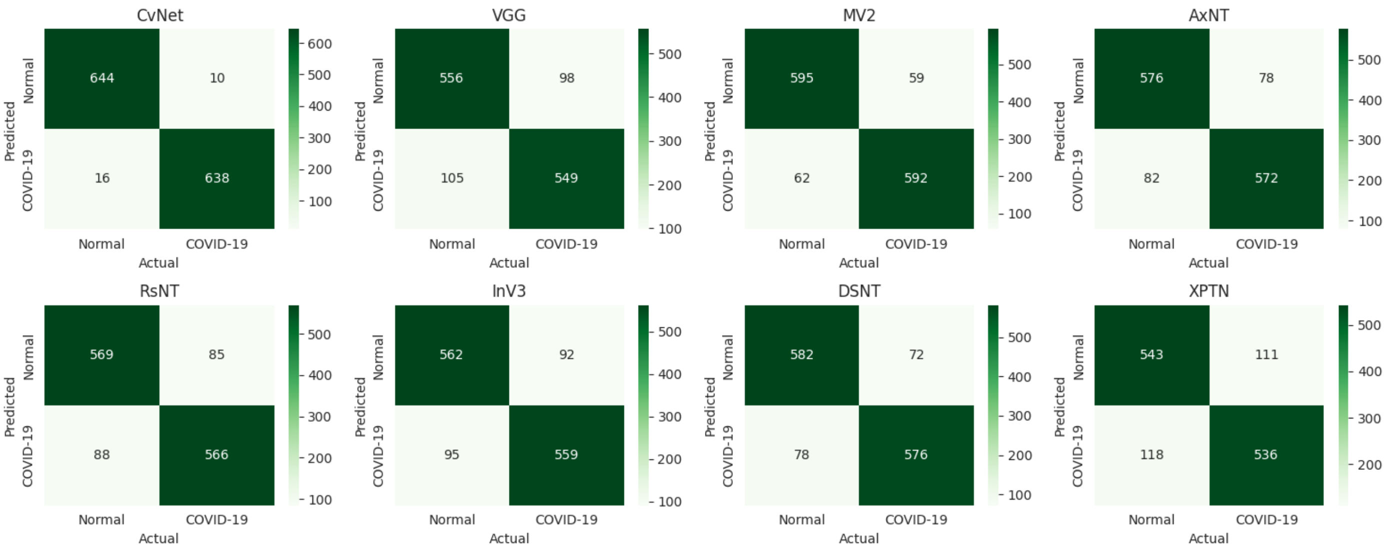


Fig. 6. Confusion matrices of deep learning models for COVID-19 vs. normal classification on dataset 1.

C. Performance for Dataset 2

From the examination of dataset 2, CVNet emerged as the most efficient classifier, obtaining an accuracy of 99.30%. Specifically, with CVNet, we observe an exceptional precision of 99.4%. Following that, MV2 obtained a score of 93.2%. Turning to the recall parameter, RsNT and InV3 exhibit impressive results, but CVNet surpasses with a significant score of 99.2%. In terms of F1-score, whereas Xception (XPTN) achieves a score of 90.10%, CVNet reaches the highest point with a score of 99.30%. For specificity, an essential statistic for distinguishing an actual negative rate, CVNet continues to dominate with 99.3%, outperforming the other models.

Other metrics, including NPV, FOR, FPR, FDR, FNR, and the Kappa Score, further emphasize the capabilities of each model. Furthermore, CvNet differentiates itself across these parameters. With an NPV of 99.3%, it greatly surpasses VGG, which settles at 89.1%. Analyzing the FOR, both VGG and InV3 score better with 10.9% and 16.3%, respectively, although CvNet excels with a low score of 0.7%. Focusing on the FPR, models such as VGG, RsNT, and XPTN exhibit increased rates. Yet again, CvNet outperforms with a minimum 0.7%. In terms of the FDR, XPTN rate of 9.7% compares strongly with CvNet only 0.6%. On the FNR, XPTN displays a higher number, while CvNet demonstrates its efficiency with a modest 0.8%. Finally, while reviewing the Kappa Score, the quality of CvNet shows effectively with 0.986, indicating a nearly perfect agreement with true labels. In comparison, XPTN obtains 0.802, with MV2 and DSNT slowing with their respective scores. Fig. 7 visually represents the model performances with a comprehensive comparison.

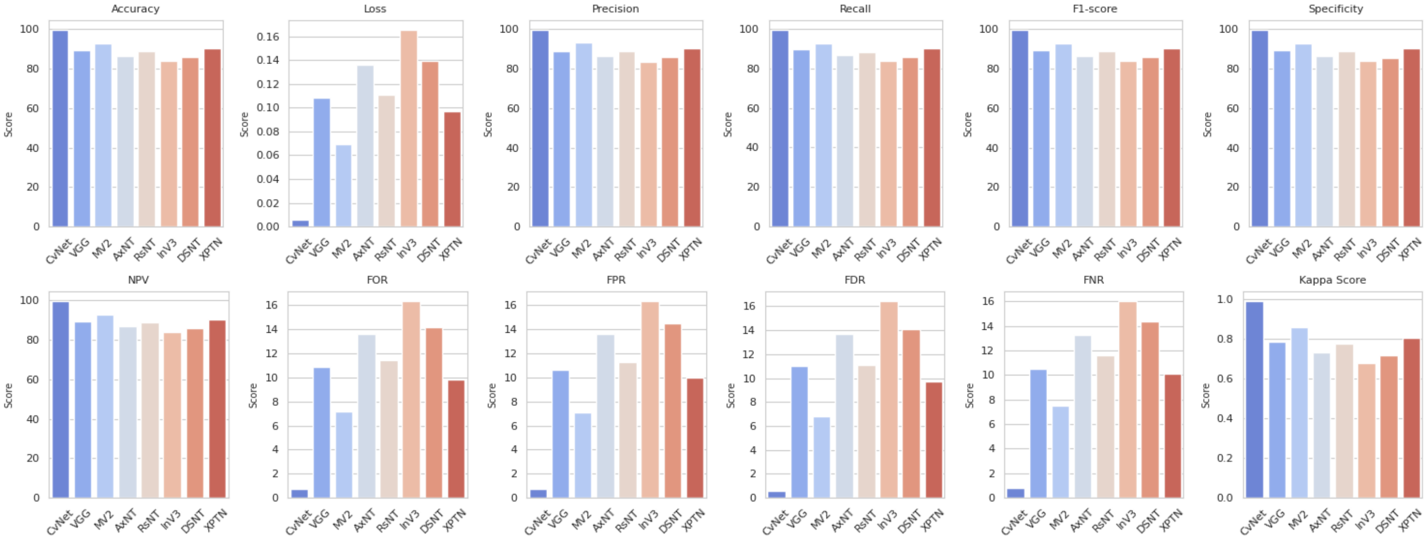


Fig. 7. Comparative performance metrics across the models on dataset 2.

Fig. 8 illustrates the confusion matrix for all of the employed models including CvNet, VGG, MV2, AxNT, RsNT, InV3, DSNT, and XPTN) in the context of identifying between COVID-19 and normal cases. Upon assessing the CvNet model, it's observed that out of the total COVID-19 cases, 495 were accurately recognized as COVID-19 (True Positives), whereas seven were misclassified as normal (False Negatives). In comparison, for the normal cases, 493 occurrences were adequately identified as normal (True Negatives), while five were mislabeled as COVID-19 (False Positives). Observing comparable patterns with different models, take the VGG model as an instance: it correctly recognized 441 occurrences of COVID-19, but mislabeled 61 as normal. Concurrently, out of the normal samples, 439 were appropriately identified, while 59 were incorrectly classed as COVID-19. Despite the variances between models, the performance measures indicate CvNet as the standout, exemplifying higher classification for dataset 2 compared to other models.

D. Performance for Dataset 3

From the assessment of our dataset 3, CvNet emerges as the best classifier with an outstanding accuracy of 98.25%. Specifically, with CvNet, we see an outstanding precision of 98.7%. Besides, MV2 achieves an accuracy of 93.8%. As for the recall, both RsNT and AxNT produced substantial results, CvNet reached out with a higher score of 98%. Moving to the F1-score, though XPTN gains a score of 91.9%, CvNet achieves its highest rating with 98.35%. In specificity, a critical parameter for determining the real negative rate, CvNet maintains its dominance with a remarkable 98.9%, surpassing the other models.

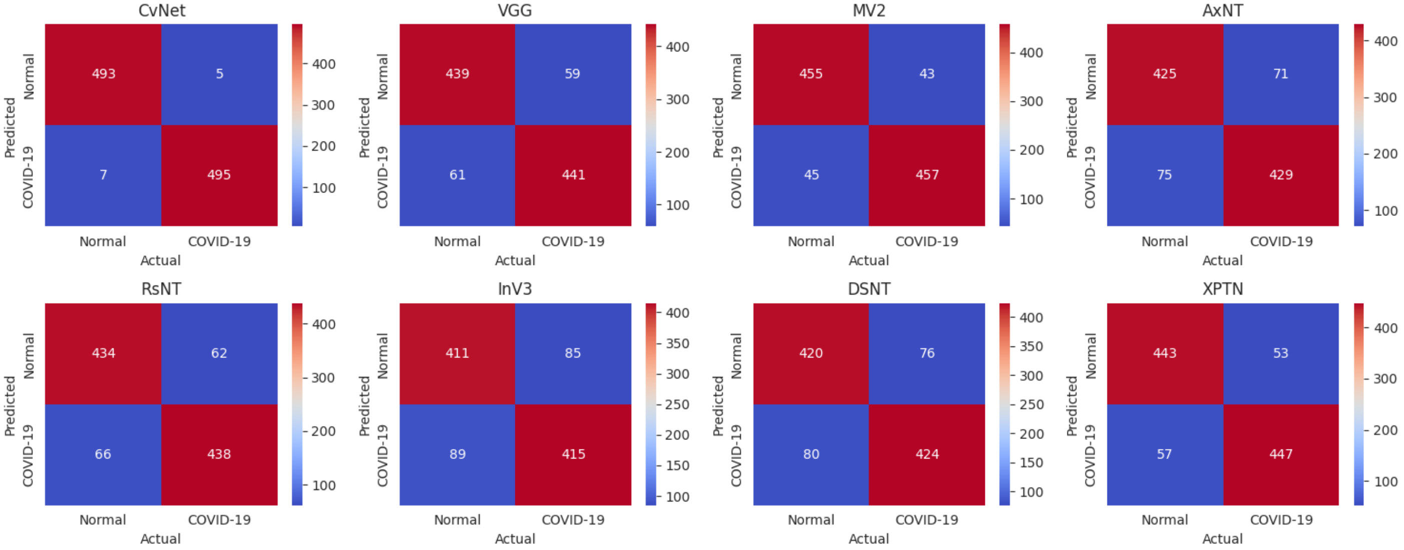


Fig. 8. Confusion matrices of deep learning models for COVID-19 vs. normal classification on dataset 2.

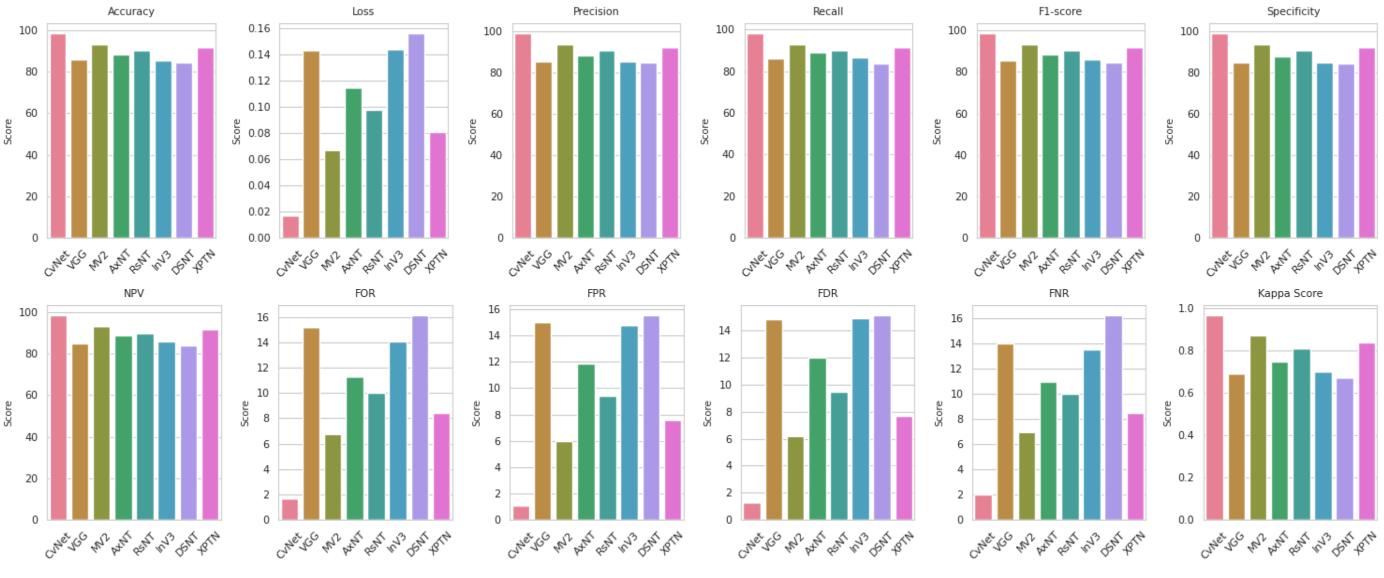


Fig. 9. Comparative performance metrics across the models on dataset 3.

We learn more about each model's effectiveness by exploring other metrics, such as NPV, FOR, FPR, FDR, FNR, and the Kappa Score. CvNet is one of them that sticks out particularly. CvNet significantly exceeds VGG, which has an NPV of 84.8%, with an NPV of 98.3%. In contrast to CvNet's admirable 1.7%, both VGG and DSNT models exhibit greater rates when analyzing FOR, 15.2% and 16.1%, respectively. The VGG, RsNT, and DSNT exhibit greater rates in terms of FPR, CvNet excels with a rate of only 1.1%. In terms of the FDR, XPTN has a respectable rate of 7.7%, whereas CvNet excels with only 1.3%. XPTN has a little higher value while monitoring FNR, however, CvNet performs better with pure 2%. Lastly, CvNet's Kappa Score of 0.965 is remarkable and perfectly matches true labels. This rating distinguishes it from other models like XPTN, MV2, and DSNT. Fig. 9, a visual representation, helps clarify these comparisons and provides an extensive overview of how well the models performed, showing their advantages and possible weaknesses.

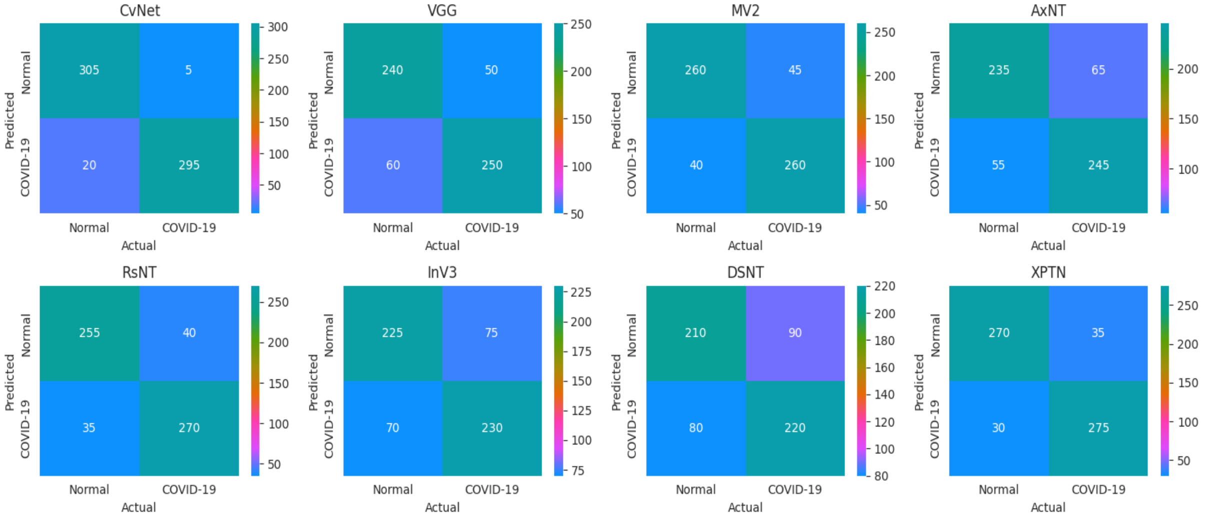


Fig. 10. Confusion matrices of deep learning models for COVID-19 vs. normal classification on dataset 3.

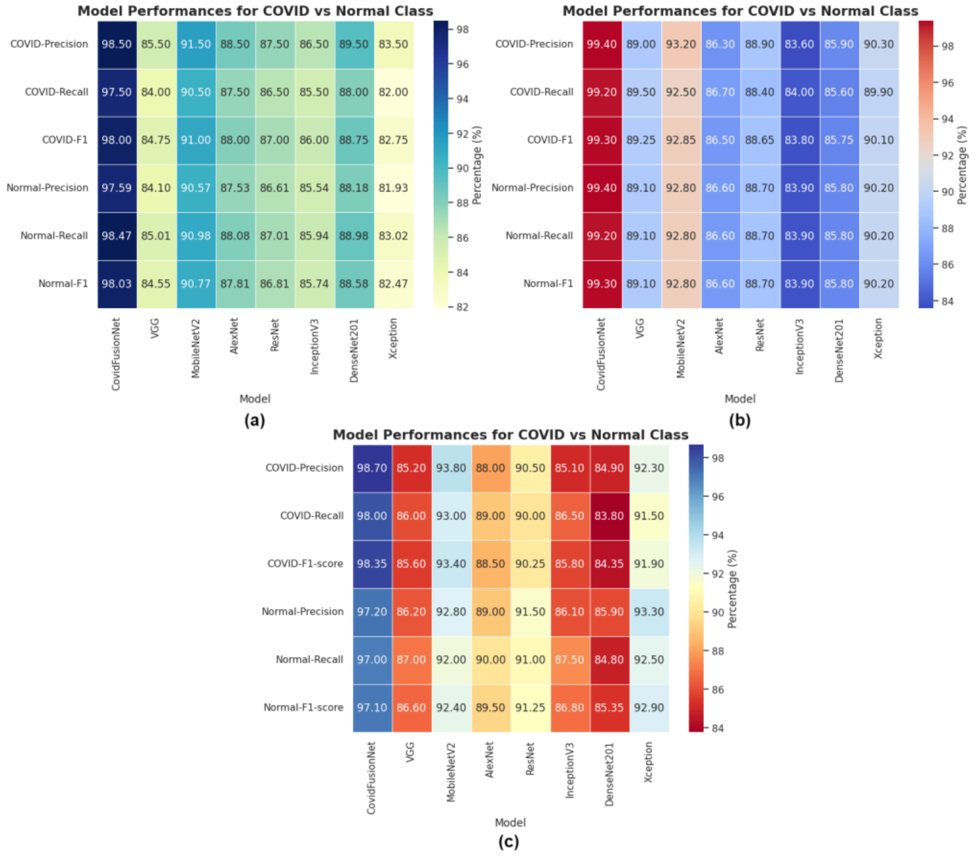


Fig. 11. Class-wise performance metrics of various models for COVID-19 detection across three datasets: (a) Dataset 1, (b) Dataset 2, and (c) Dataset 3

Fig. 10 exhibits the confusion matrices for all the models evaluated, including CvNet, VGG, MV2, AxNT, RsNT, InV3, DSNT, and XPTN, in their attempts to identify COVID-19 and normal cases. A detailed review of the CvNet model indicates that of the total supposed COVID-19 cases, 295 were accurately recognized as COVID-19 (True Positives), whereas 20 were erroneously tagged as normal (False Negatives). Conversely, among the samples identified as normal, 305 were properly marked as normal (True Negatives), with 5 being mistakenly indicated as COVID-19 (False Positives). VGG model properly discovered 250 instances of COVID-19, but wrongly categorized 60 as normal. However, out of the normal samples, 240 were correctly recognized, while 50 were incorrectly categorized as COVID-19. Moreover, CvNet distinguishes itself as being better by classifying Dataset 3 with more accuracy when compared to the other models.

E. Classwise Performance Analysis

Fig. 11(a) exhibiting Dataset 1 indicates that CovidFusionNet produced outstanding accuracy for the COVID class (98.5%). For the COVID class, CovidFusionNet achieved a highest F1-score (98%) whereas ResNet closely followed with an F1-score of 87%. For the Normal class, CovidFusionNet scored the best accuracy value (97.59%) and its recall was equally effective at 98.47%. This implies that for the categories COVID and Normal, while CovidFusionNet outperforms in one measure, other models such as ResNet or AlexNet tend to follow closely, showing their complementing potential. Fig. 11(b) depicting Dataset 2 demonstrates that CovidFusionNet once again came out with the highest accuracy for the COVID class (99.4%). In terms of the COVID class, CovidFusionNet retained the top F1-score (99.30%). For the Normal class, both accuracy and recall values were generally stable across the models, with CovidFusionNet winning once again with an F1-score of 99.3%. The superior performance of CovidFusionNet across these measures reinforces its great competence in discriminating between both classes. Fig. 11(c) from Dataset 3 highlights that CovidFusionNet continues its trend with outstanding accuracy for the COVID class (98.7%). MobileNetV2 obtained an excellent F1-score (93.4%) for the COVID class, whereas for the Normal class, Xception performed amazingly with the top F1-score (92.9%).

IV. Discussion

The CovidFusionNet is specifically designed for high-resolution medical imaging, with a focus on X-ray images. It showcases its capabilities by combining several Convolutional Neural Networks (CNNs) with Wavelet Transforms, which are essential for precise identification of patterns and reliable diagnosis of COVID-19. The oversampling strategy employed by this system enhances its ability to effectively handle datasets with class imbalances, which are often seen in medical data. As a result, this system is especially relevant in the present healthcare scene. In addition, CovidFusionNet's ability to do multi-resolution image analysis, utilizing both Continuous and Discrete Wavelet Transform, enables the detection of anomalies at various scales. Furthermore, integrating Adaptive Histogram Equalization with Wavelet Transforms for image improvement is particularly effective for datasets requiring precise feature retention and noise reduction. Although it demonstrates exceptional performance in these domains, its utilization with other data formats, such as MRI or CT scans, may necessitate appropriate modifications to achieve ideal outcomes. In consideration of this recognition, further study will prioritize the extension of CovidFusionNet's utilization to a range of data formats, thereby thoroughly assessing its efficacy in different clinical environments. This study aims to improve the effectiveness and expand the range of applications of the model in medical imaging, thereby transforming it into a flexible instrument in the advancing field of medical diagnostics. We have compared the performance of our proposed model with state-of-the-art methods, and the comparative results are represented in Table III.

TABLE III. Comparing Proposed Method Performance with State-Of-The-Art Methods

Reference	Dataset	Methodology	Accuracy
[14]	X-ray	CNN	91%
[15]	X-ray	Covid-Aid	87%
[16]	X-ray	CNN +GRU	93%
[17]	X-ray	STM-RENet	96.53%
[18]	X-ray	InceptionResNetV2 + Xception	95.78%
[20]	X-ray	ResNet50 + SVM	94.7%
[21]	X-ray	ResNet152V2+GRU	93.37%
[22]	X-ray	VGG19 + CNN	96.48%
Proposed	X-ray	CovidFusionNet	Dataset 1 (98.02%)
			Dataset 2 (99.30%)
			Dataset 3 (98.25%)

V. Conclusion

In this study, we introduced the CovidFusionNet, a cutting-edge CNN model based on fusion and optimized for accurately classifying COVID-19 X-ray images across three distinct datasets. We improved image clarity and detail preservation by combining features from seven pre-trained convolutional neural networks, incorporating the Continuous and Discrete Wavelet Transform, and using an innovative enhancement technique that combines Adaptive Histogram Equalization and Wavelet Transforms (HAWIE). Our model outperformed seven well-known pre-trained models in accuracy and consistency when combined with an oversampling strategy to solve the class imbalance. This work advances the field of image-based COVID-19 diagnosis by providing a tool ready for clinical use and pointing out potential directions for further investigation into the effects of image quality on detection effectiveness.

Data Availability Statement: Dataset 1 can be found at https://www.kaggle.com/datasets/unaissait/curated-chest-xray-image-dataset-for-covid19 (Accessed on 1 September 2023); Dataset 2 can be found at https://www.kaggle.com/datasets/amanullahasraf/covid19-pneumonia-normal-chest-xray-pa-dataset, (Accessed on 2 September 2023); Dataset 3 can be found at https://www.kaggle.com/datasets/jtiptj/chest-xray-pneumoniacovid19tuberculosis, (Accessed on 2 September 2023).

References

[1] R. Huang, M. Liu, and Y. Ding, ‘‘Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis,’’ J. Infection Developing Countries, vol. 14, no. 3, pp. 246–253, Mar. 2020.

[2] A. C. Cunningham, H. P. Goh, and D. Koh, ‘‘Treatment of COVID-19: Old tricks for new challenges,’’ Crit. Care, vol. 24, no. 1, p. 91, Dec. 2020.

[3] Z. Wang, Y. Xiao, Y. Li, J. Zhang, F. Lu, M. Hou, and X. Liu, ‘‘Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays,’’ Pattern Recognit., vol. 110, Feb. 2021, Art. no. 107613.

[4] D. Cucinotta and M. Vanelli, ‘‘WHO declares COVID-19 a pandemic,’’ Acta Bio Medica, Atenei Parmensis, vol. 91, no. 1, p. 157, 2020.

[5] M. Ndiaye, S. S. Oyewobi, A. M. Abu-Mahfouz, G. P. Hancke, A. M. Kurien, and K. Djouani, ‘‘IoT in the wake of COVID-19: A survey on contributions, challenges and evolution,’’ IEEE Access, vol. 8, pp. 186821–186839, 2020.

[6] H. A. Rothan and S. N. Byrareddy, ‘‘The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak,’’ J. Autoimmunity, vol. 109, May 2020, Art. no. 102433.

[7] S. Hu, Y. Gao, Z. Niu, Y. Jiang, L. Li, X. Xiao, M. Wang, E. F. Fang, W. Menpes-Smith, J. Xia, H. Ye, and G. Yang, ‘‘Weakly supervised deep learning for COVID-19 infection detection and classification from CT images,’’ IEEE Access, vol. 8, pp. 118869–118883, 2020.

[8] C. Menni, A. Valdes, M. B. Freydin, S. Ganesh, J. E.-S. Moustafa, A. Visconti, P. Hysi, R. C. Bowyer, M. Mangino, M. Falchi, and J. Wolf, ‘‘Loss of smell and taste in combination with other symptoms is a strong predictor of COVID-19 infection,’’ MedRxiv, 2020.

[9] L. Goyal and N. Arora, ‘‘Deep transfer learning approach for detection of COVID-19 from chest X-ray images,’’ Int. J. Comput. Appl., vol. 975, p. 8887.

[10] M. Awais, M. Raza, N. Singh, K. Bashir, U. Manzoor, S. U. Islam, and J. J. P. C. Rodrigues, ‘‘LSTM based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19,’’ IEEE Internet Things J., early access, Dec. 10, 2020, doi: 10.1109/JIOT.2020.3044031.

[11] I. Kokkinakis, K. Selby, B. Favrat, B. Genton, and J. Cornuz, ‘‘COVID-19 diagnosis: Clinical recommendations and performance of nasopharyngeal swab-PCR,’’ Revue Medicale Suisse, vol. 16, no. 689, pp. 699–701, 2020.

[12] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, J. R. Weir-McCall, Z. Teng, E. Gkrania-Klotsas, J. H. F. Rudd, E. Sala, and C.-B. Schönlieb, ‘‘Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans,’’ Nature Mach. Intell., vol. 3, no. 3, pp. 199–217, Mar. 2021.

[13] C.-C. F. Tam, K. S. Cheung, S. Lam, A. Wong, A. Yung, M. Sze, Y.-M. Lam, C. Chan, T. C. Tsang, M. Tsui, and H. F. Tse, ‘‘Impact of coronavirus disease 2019 (COVID-19) outbreak on ST-segment–elevation myocardial infarction care in Hong Kong, China,’’ Circulat., Cardiovascular Qual. Outcomes, vol. 13, no. 4, 2020, Art. no. e006631

[14] Gao, T., & Wang, G. (2020). Chest X-ray image analysis and classification for COVID-19 pneumonia detection using Deep CNN. medRxiv, 2020-08.

[15] Singh, S., Sapra, P., Garg, A., & Vishwakarma, D. K. (2021, April). CNN based Covid-aid: Covid 19 Detection using Chest X-ray. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1791-1797). IEEE.

[16] Shah, P. M., Ullah, F., Shah, D., Gani, A., Maple, C., Wang, Y., Abrar, M., & Islam, S. U. (2021). Deep GRU-CNN model for COVID-19 detection from chest X-rays data. Ieee Access, 10, 35094-35105.

[17] Khan, S. H., Sohail, A., Khan, A., & Lee, Y. S. (2022). COVID-19 detection in chest X-ray images using a new channel boosted CNN. Diagnostics, 12(2), 267.

[18] Gayathri, J. L., Abraham, B., Sujarani, M. S., & Nair, M. S. (2022). A computer-aided diagnosis system for the classification of COVID-19 and non-COVID-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. Computers in biology and medicine, 141, 105134.

[19] Banerjee, A., Sarkar, A., Roy, S., Singh, P. K., & Sarkar, R. (2022). COVID-19 chest X-ray detection through blending ensemble of CNN snapshots. Biomedical Signal Processing and Control, 78, 104000.

[20] Ismael, A. M., & Şengür, A. (2021). Deep learning approaches for COVID-19 detection based on chest X-ray images. Expert Systems with Applications, 164, 114054.

[21] Kanjanasurat, I., Tenghongsakul, K., Purahong, B., & Lasakul, A. (2023). CNN–RNN Network Integration for the Diagnosis of COVID-19 Using Chest X-ray and CT Images. Sensors, 23(3), 1356.

[22] Alshmrani, G. M. M., Ni, Q., Jiang, R., Pervaiz, H., & Elshennawy, N. M. (2023). A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. Alexandria Engineering Journal, 64, 923-935.

[23] Kuzinkovas, D., & Clement, S. (2023). The detection of covid-19 in chest x-rays using ensemble cnn techniques. Information, 14(7), 370.

[24] Hafeez, U., Umer, M., Hameed, A., Mustafa, H., Sohaib, A., Nappi, M., & Madni, H. A. (2023). A CNN based coronavirus disease prediction system for chest X-rays. Journal of Ambient Intelligence and Humanized Computing, 14(10), 13179-13193.

[25] Shamrat, F. J. M., Azam, S., Karim, A., Ahmed, K., Bui, F. M., & De Boer, F. (2023). High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. Computers in Biology and Medicine, 155, 106646.

[26] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.

[28] Krizhevsky, A. (2014). One weird trick for parallelizing convolutional neural networks. arXiv preprint arXiv:1404.5997.

[29] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[30] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[31] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[32] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[33] Akter, S., Shamrat, F. J. M., Chakraborty, S., Karim, A., & Azam, S. (2021). COVID-19 detection using deep learning algorithm on chest X-ray images. Biology, 10(11), 1174

Implementation of a Convolutional Neural Network (CNN)-based Object Detection Approach for Smart Surveillance Applications

Weiguo Ni*
School of UAV, Guangzhou Civil Aviation College, Guangzhou 510000, Guangdong, China

***Abstract*—In the realm of smart surveillance systems, a fundamental technique for tracking and evaluating consumer behavior is object detection through video surveillance. While existing research underscores object detection through deep learning techniques, a notable gap exists in adapting these methods to effectively capture and recognize small, intricate objects. This study addresses this gap by introducing a customized methodology tailored to meet the nuanced requirements of accurate and lightweight detection for small objects, especially in scenarios prone to visual complexity and object similarity challenges. The primary objective is to furnish a vision-based object identification method designed for surveillance applications in smart stores, with a particular focus on locating jewelry objects. To achieve this, a Convolutional Neural Network (CNN)-based object detector utilizing YOLOv7 is employed for precise object detection and location extraction. The YOLOv7 network undergoes rigorous training and verification on a unique dataset specifically curated for this purpose. Experimental results affirm the efficacy of the proposed object identification method, demonstrating its capacity to detect items relevant to smart surveillance applications.**

Keywords—Smart surveillance; lightweight object detection; YOLOv7; small object recognition; vision-based identification

I. Introduction

Cameras and image sensors are frequently deployed in smart surveillance systems so that automated object identification techniques may be used to automatically detect and identify various objects in smart environment analysis [1, 2]. Such automatic object recognition techniques often need sophisticated image/data processing tools and algorithms [3]. As a result, developing low-complexity automated object identification algorithms for use in urban surveillance applications becomes crucial [4, 5].

For computer vision applications, deep learning-based approaches, including Convolutional Neural Networks (CNNs), are among the finest solutions [6, 7]. Applications like object categorization and image segmentation have made significant strides thanks to CNNs [8]. Additionally, CNNs contain convolution layers that handle feature extraction; they are resilient to shifts and distortions in the image; they use less memory; training is simpler; and, as a result of the fewer parameters, they are better and quicker [9].

Object detection and monitoring in IoT Smart Shop Surveillance Systems have witnessed significant advancements in recent years. Current technologies utilize a combination of computer vision, IoT devices, and machine learning algorithms to enhance security, customer experience, and operational efficiency in retail environments. The integration of IoT devices enables real-time data collection from various sensors and cameras while object detection algorithms process this data to identify and track objects of interest.

In previous studies, various methods have been explored for object detection and monitoring in IoT Smart Shops [10, 11]. Traditional approaches, such as handcrafted features and rule-based algorithms, have limitations in handling complex and diverse scenarios. However, deep learning-based methods, particularly the CNNs, have gained immense popularity [12, 13]. Deep learning models can automatically learn and extract relevant features from raw data, making them capable of handling complex object detection tasks. The ability of deep learning models to handle large-scale datasets and their superior performance in terms of accuracy have attracted researchers to explore and develop new approaches based on these techniques.

Despite the advancements, there are still some limitations and research gaps in the field. One major challenge is the requirement for low computational costs and high accuracy rates [14, 15]. Many IoT devices have limited processing power and memory, making it necessary to develop lightweight deep-learning algorithms that can achieve high accuracy while maintaining computational efficiency. Additionally, the lack of publicly available datasets specifically designed for IoT Smart Shop Surveillance Systems poses another challenge for researchers.

To address these limitations, researchers have focused on developing lightweight deep learning models and utilizing algorithms like YOLO (You Only Look Once) for efficient object detection. YOLO-based algorithms offer real-time object detection capabilities with relatively low computational requirements [16, 17]. Using custom datasets, researchers can train these models on specific Smart Shop Surveillance scenarios, encompassing various objects and environmental factors.

In this study, 1170 images were collected for our custom dataset from the Internet and our capturing webcam-based process, and the image augmentation process in our custom dataset preparation. The dataset is used for training and evaluating a model based on the YOLOv7 network. This model has generated a weight and is used to perform object recognition using the YOLOv7 model on our custom dataset.

This study introduces novel contributions in the field of computer vision and deep learning for IoT Smart Shop Surveillance Systems. It innovates by developing a lightweight deep learning model tailored to the limited computational resources of IoT environments. The creation of a custom dataset, incorporating 1170 images with meticulous preparation, stands out as a unique aspect, emphasizing the study's commitment to robust methodology. The adoption of the YOLOv7 network architecture for object recognition further highlights the innovative application of state-of-the-art technologies to address surveillance challenges in a Smart Shop context.

In terms of research contributions, three potential areas of focus are identified. Firstly, the development of a lightweight deep learning model tailored for IoT Smart Shop Surveillance Systems, considering the low computational resources available. Secondly, the creation of a custom dataset that represents realistic scenarios encountered in Smart Shop Surveillance. This dataset can enable the training and evaluation of the proposed model. Finally, conducting thorough experimental evaluations to assess the performance of the model in terms of accuracy, real-time detection, and computational efficiency. By addressing the research gap through the proposed lightweight deep learning model, custom dataset, and rigorous performance evaluations, researchers can contribute to advancing object detection and monitoring in IoT Smart Shop Surveillance Systems. The outcome of this research can lead to improved security, customer experience, and operational efficiency in retail environments, promoting the widespread adoption of IoT-based surveillance systems in the retail industry.

The structure of this paper is as follows; Section I presents the introduction. The proposed approach discusses in Section II. Section III involves experimental results, and Section IV concludes the paper.

II. Related Works

This section reviews the related works that are focused on object detection in video-based surveillance systems.

Mneymneh et al. [18] introduced a vision-based framework for intelligent monitoring of hardhat wearing on construction sites. The framework utilizes computer vision techniques to detect and track the presence of hardhats on individuals within the construction site environment. It involves stages of image acquisition, pre-processing, detection, and monitoring to identify and track individuals wearing hardhats. The study's limitations include reliance on a specific color-based segmentation approach, vulnerability to challenging lighting conditions, and the absence of exploration of other safety equipment detection. Addressing these limitations can enhance the framework's accuracy and broaden its applicability in ensuring compliance with safety regulations on construction sites.

Lu et al. [19] presented a real-time object detection algorithm for video. The algorithm utilizes a combination of deep learning techniques, including Convolutional Neural Networks (CNNs) and feature extraction methods, to detect objects in video frames. The proposed algorithm achieves high detection accuracy and real-time performance by optimizing the architecture and leveraging parallel processing capabilities. However, the limitation of the study is that the algorithm's performance may be affected by complex scenes with occlusions or high object density, which can lead to missed detections or false positives. Further research could focus on improving the algorithm's robustness in challenging video scenarios to enhance its overall effectiveness in real-world applications.

The authors in [20] presented a methodology based on deep learning for object detection in video surveillance, specifically focusing on the identification of small objects that are handled similarly. The proposed methodology utilizes binary classifiers and leverages deep learning techniques to achieve accurate object detection. The approach demonstrates promising results in detecting small objects in challenging video surveillance scenarios. However, a limitation of the study is that the proposed methodology may face challenges when dealing with highly cluttered scenes or objects that have similar visual characteristics but different semantic meanings. Further research could explore techniques to address these limitations and improve the methodology's robustness in handling complex scenarios, ultimately enhancing its applicability in video surveillance applications.

Alrowais et al. [21] developed a deep transfer learning-enabled intelligent object detection approach for crowd density analysis in video surveillance systems. The proposed method utilizes deep learning techniques and transfers learning to detect and analyze crowd density in video footage. By leveraging pre-trained models and fine-tuning them on specific crowd density datasets, the approach achieves accurate object detection and density analysis. The results demonstrate the effectiveness of the method in crowd density estimation. However, a limitation of the study is the reliance on pre-trained models that may not fully capture the diverse range of crowd dynamics and behaviors. Further research could focus on developing customized models or incorporating additional data augmentation techniques to improve the algorithm's performance and robustness in capturing different crowd scenarios.

According to review of previous studies, Addressing the research challenge of achieving high accuracy and lightweight object detection, particularly for small objects like jewelry, requires a tailored approach. While existing studies focus on object detection using deep learning techniques, adapting these methods to effectively capture and recognize small, intricate objects remains a gap. Current methodologies may struggle in cluttered scenes or with objects sharing similar visual characteristics.

III. Proposed Approach

This study selects jewelry objects and implements the approach for these kinds of objects involving rings and earnings. We suggest a detection method based on YOLOv7 to create a model that could recognize jewelry [9].

A. YOLO

The YOLO model, which stands for "You Only Look Once," is one stage object detector [17]. YOLO predicts the positions of the bounding boxes and the classes of the bounding boxes [19]. Objectness of the bounding boxes after feature-stripping image frames through a backbone, combining and blending features in the neck, and objectness prediction in the head of the network [22]. To arrive at its ultimate forecast, YOLO employs post-processing using NMS [23]. Fig. 1 illustrates the basic concept of YOLO network architecture.

B. Dataset

Our dataset consists of two various categories of jewelry (earrings and rings), as mentioned earlier. The dataset includes images from the Internet and our captured images in the real environment. We gathered webcam photos from two fixed cameras that were placed in various places. We chose images with a variety of model kinds, sizes, resolutions, orientations, and sample counts in each picture. One thousand one hundred seventy example photos of two types of jewelry pieces in various orientations, rotations, and scales are included in our unique dataset. Some sample images from our dataset are shown in Fig. 2.

C. Training and Testing

It is usually a good idea to start with a model that has already been trained using extremely big datasets and then utilize the model's weights to train an object detector [24]. Even if the learned weights don't contain the items needed for this specific experiment, to deal with the training process in the YOLO network, initial weights are taken from a pretrained model that includes weights from the known dataset [25]. In this study, we use an initial weight that trained a model from the COCO dataset.

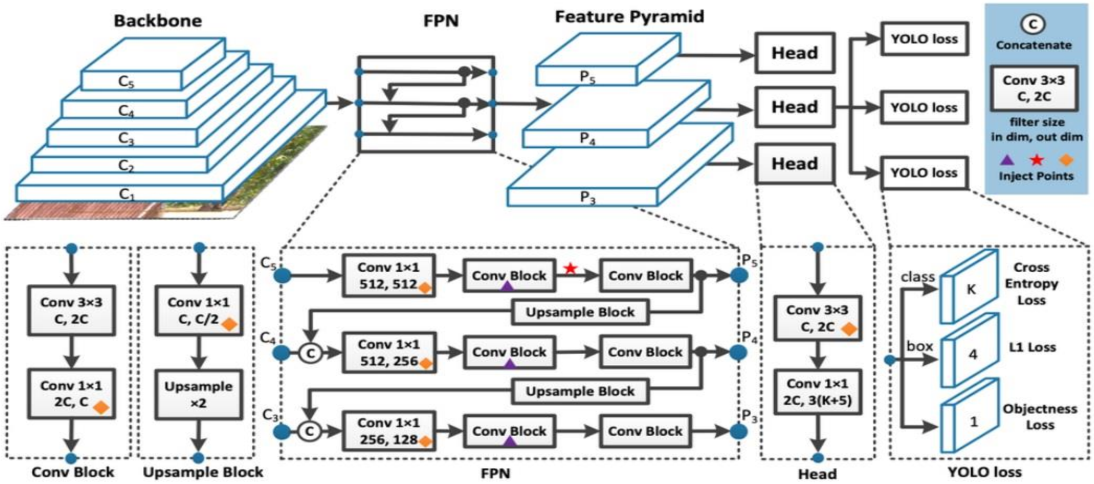


Fig. 1. YOLO network architecture [23].

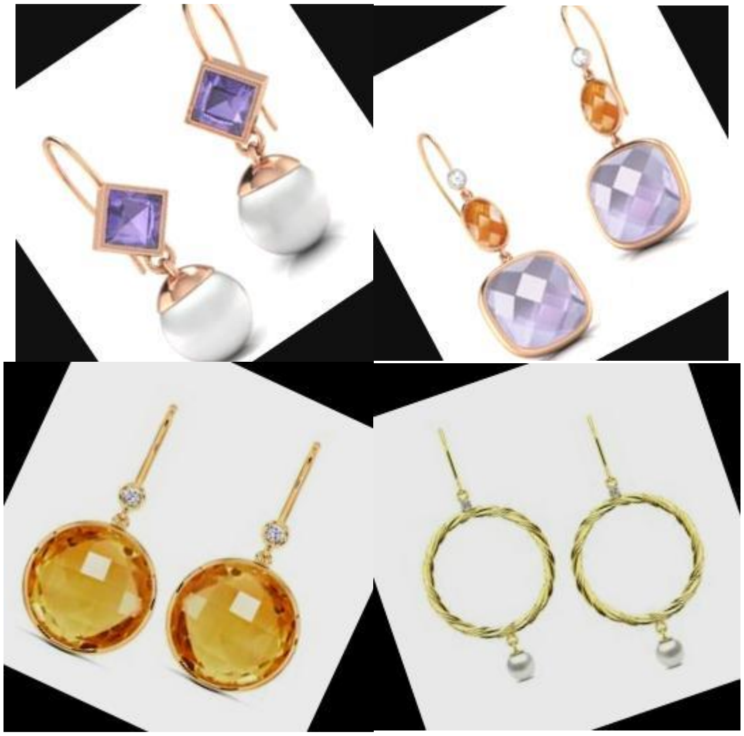


Fig. 2. Some image samples from the dataset.

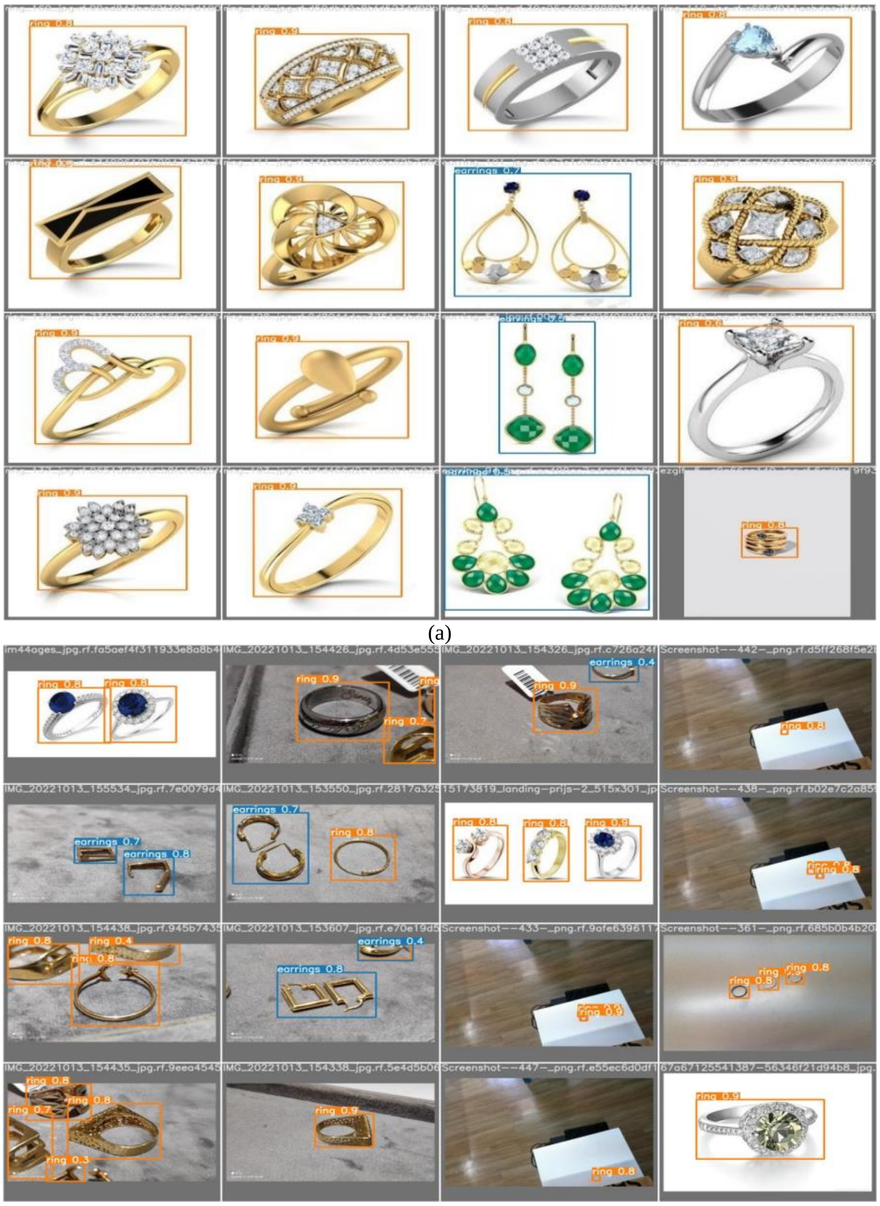
IV. Experimental Results

In this section, we present the experiment's details, and then we show the training results using pretraining weights and compare the three models of YOLOv7. Fig. 3 shows the result of the implementation of our proposed approach.

In the following, some analyses are presented to justify why the YOLOv7 can be presented accurate results and has superiority to apply in real-time requirements. To prove this superiority, some visual representations in graphs are illustrated. Using these graphs, a comparison of performance results is shown to visually demonstrate the superiority justifications. Inspired by [26], the graph depicts a comparison of YOLOv7, YOLOv4, PPYOLOE, YOLOX, YOLOR, YOLOv5, and Transformers object detectors. The X-axis represents the Inference Time, indicating the time taken for object detection, while the Y-axis represents the average precision (%) of the detectors.

Average precision (AP) is a commonly used metric in object detection that measures the accuracy of a model in localizing and classifying objects. It calculates the precision at various recall levels, considering the trade-off between precision and recall. Higher AP values indicate better performance and accuracy in object detection.

Inference Time refers to the time taken by the model to perform object detection on a given input. It reflects the efficiency and speed of the algorithm in processing frames or images in real-time applications. Lower inference times are desirable for real-time object detection scenarios. Fig. 4 presents the comparison of object detectors [26].



(a)

(b)

Fig. 3. Experimental results (a) and (b).

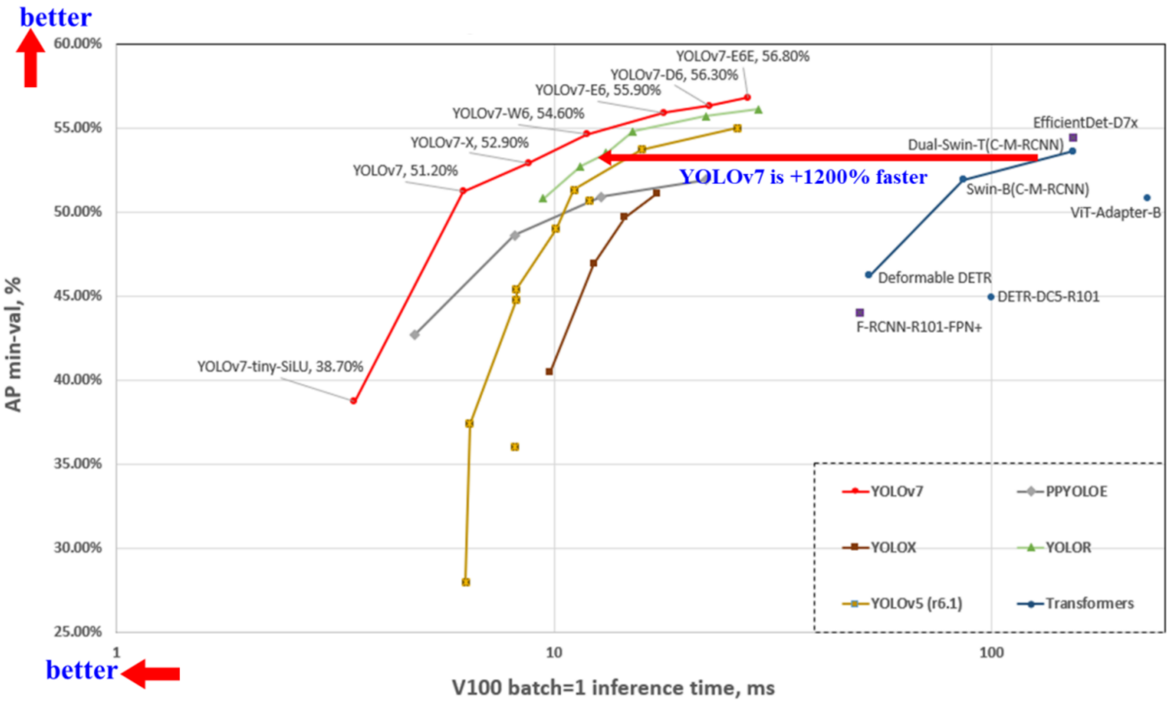


Fig. 4. Comparison of object detectors [26].

As shown in Fig. 4, it can be observed that YOLOv7 outperforms the other object detectors in terms of precision rate. At similar inference times, YOLOv7 consistently achieves higher average precision compared to PPYOLOE, YOLOX, YOLOR, YOLOv5, and Transformers. This suggests that YOLOv7 demonstrates superior accuracy in detecting and recognizing objects across various scenarios.

The YOLOv7 utilizes an efficient single-pass detection pipeline that eliminates the need for time-consuming region proposal techniques. This allows YOLOv7 to process images and videos in real-time without compromising accuracy. Other detectors may achieve lower inference times but often at the expense of reduced precision. Furthermore, YOLOv7 incorporates advanced training strategies, including data augmentation techniques and optimization methods like focal loss and learning rate scheduling. These strategies enhance the model's ability to generalize and accurately detect objects under diverse conditions, contributing to its superior precision rate.

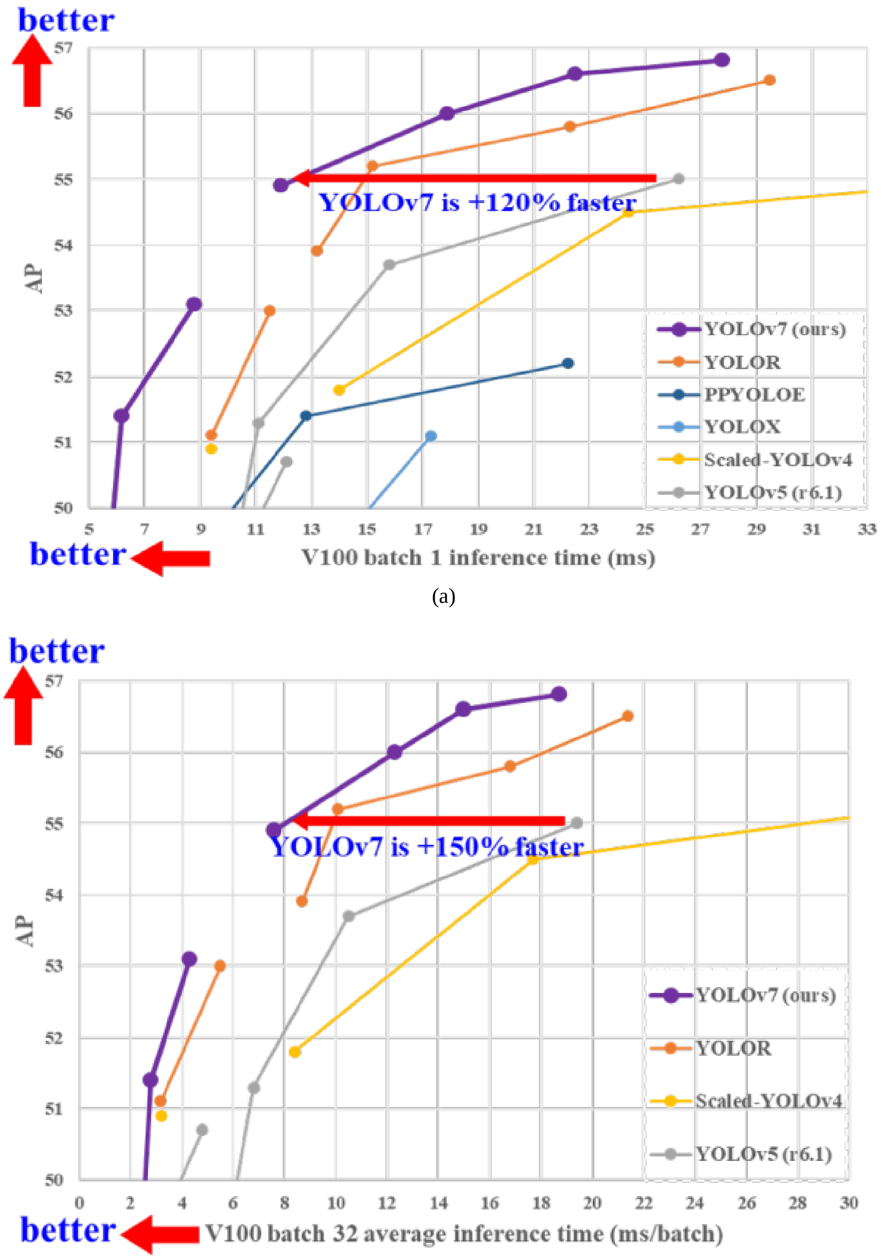
As a result, the graph demonstrates that YOLOv7 outperforms PPYOLOE, YOLOX, YOLOR, YOLOv5, and Transformers in terms of precision rate. The advanced architecture, efficient detection pipeline, and advanced training strategies employed by YOLOv7 contribute to its effectiveness and accuracy in object detection.

In the following, performance analysis and comparison of AP are presented to justify why the YOLOv7 is better than other object detector algorithms. Fig. 4 demonstrates the comparison of object detectors in real-time [26].

Fig. 5 illustrates the comparison of object deters in real-time conditions based on the presented comparison graphs. The graph illustrates a comparison of real-time object detection algorithms, including YOLOv7, PPYOLOE, YOLOX, YOLOR, Scaled-YOLOv4, YOLOv5, and Transformers. The X-axis represents the Inference Time, indicating the time taken for object detection, while the Y-axis represents the average precision (AP) of the detectors.

As shown in Fig. 4, it can be observed that YOLOv7 performs better than the other object detection algorithms in terms of precision rate while maintaining real-time capabilities. YOLOv7 achieves higher AP values at similar inference times compared to PPYOLOE, YOLOX, YOLOR, Scaled-YOLOv4, YOLOv5, and Transformers. This indicates that YOLOv7 provides more accurate object detection results. Furthermore, YOLOv7 demonstrates its effectiveness in real-time conditions by achieving low inference times while maintaining high precision. It strikes a balance between accuracy and speed, making it suitable for real-time applications where accuracy and efficiency are crucial.

As depicted in Fig. 4 and Fig. 5, the comparative analysis highlights the exceptional performance of YOLOv7 in the realm of object detection. The YOLOv7 stands out by demonstrating superior precision rates while seamlessly maintaining real-time capabilities. The discernible advantage lies in its ability to achieve higher Average Precision (AP) values when compared to a spectrum of other prominent object detection algorithms, including PPYOLOE, YOLOX, YOLOR, Scaled-YOLOv4, YOLOv5, and Transformers. This distinction is particularly noteworthy as it underscores YOLOv7's efficacy in delivering heightened accuracy without compromising on the efficiency required for real-time applications. The observed performance superiority positions YOLOv7 as a compelling choice for tasks demanding both precision and prompt response, further solidifying its standing as a leading solution in the field of object detection algorithms.



(b)

Fig. 5. Comparison of object detectors in real-time.

Finally, the graph demonstrates that YOLOv7 outperforms PPYOLOE, YOLOX, YOLOR, Scaled-YOLOv4, YOLOv5, and Transformers in terms of precision rate while maintaining real-time capabilities. Its advanced architecture, coupled with efficient inference times, makes YOLOv4 an effective and efficient choice for real-time object detection applications.

V. Conclusion

In this paper, a robust object identification technique based on YOLOv7 is developed for applications in smart surveillance. The first step involves the meticulous preparation of a custom dataset, with labels configured to adhere to the YOLOv7 format. The chosen technique demonstrates remarkable accuracy in identifying and categorizing jewelry objects, specifically rings and money. The network efficiently captures coordinates of resulting bounding boxes, enabling precise object identification within frames. While the current study focuses on training YOLOv7 for two types of jewelry, future research directions could involve expanding the model to encompass a broader spectrum of jewelry classes. This extension would enhance the model's versatility and applicability in diverse contexts within the realm of smart surveillance. Additionally, exploring real-time applications of the YOLOv7-based methodology remains an open problem, presenting an avenue for further investigation into its efficiency and effectiveness in dynamic surveillance scenarios. Moreover, investigating strategies to improve the model's adaptability to varying lighting conditions and complex backgrounds stands as a potential area for future research, contributing to the refinement of its performance in real-world surveillance applications.

ACKNOWLEDGMENT

This work was supported by Youth Innovation Talent Project of Department of Education of Guangdong Provincial (2023)-Uav Fi intelligent monitoring system based on deep learning. (Item Number：2023KQNCX158)

REFERENCES

[1] M. Mukherjee, I. Adhikary, S. Mondal, A. K. Mondal, M. Pundir, and V. Chowdary, "A vision of IoT: applications, challenges, and opportunities with Dehradun perspective," in Proceeding of international conference on intelligent communication, control and devices, 2017: Springer, pp. 553-559.

[2] G. T. S. Ho, Y. P. Tsang, C. H. Wu, W. H. Wong, and K. L. Choy, "A computer vision-based roadside occupation surveillance system for intelligent transport in smart cities," Sensors, vol. 19, no. 8, p. 1796, 2019.

[3] I. Saradopoulos, I. Potamitis, S. Ntalampiras, A. I. Konstantaras, and E. N. Antonidakis, "Edge Computing for Vision-Based, Urban-Insects Traps in the Context of Smart Cities," Sensors, vol. 22, no. 5, p. 2006, 2022.

[4] L. Hu and Q. Ni, "IoT-driven automated object detection algorithm for urban surveillance systems in smart cities," IEEE Internet of Things Journal, vol. 5, no. 2, pp. 747-754, 2017.

[5] M. J. Akhtar et al., "A Robust Framework for Object Detection in a Traffic Surveillance System," Electronics, vol. 11, no. 21, p. 3425, 2022.

[6] K. Bjerge, H. M. Mann, and T. T. Høye, "Real - time insect tracking and monitoring with computer vision and deep learning," Remote Sensing in Ecology and Conservation, vol. 8, no. 3, pp. 315-327, 2022.

[7] A. Aghamohammadi, M. C. Ang, E. A. Sundararajan, K. W. Ng, M. Mogharrebi, and S. Y. Banihashem, "Correction: A parallel spatiotemporal saliency and discriminative online learning method for visual target tracking in aerial videos," Plos one, vol. 13, no. 3, p. e0195418, 2018.

[8] J. Du, "Understanding of object detection based on CNN family and YOLO," in Journal of Physics: Conference Series, 2018, vol. 1004: IOP Publishing, p. 012029.

[9] M. Hatab, H. Malekmohamadi, and A. Amira, "Surface defect detection using YOLO network," in Proceedings of SAI Intelligent Systems Conference, 2020: Springer, pp. 505-515.

[10] J. Xu et al., "Design of smart unstaffed retail shop based on IoT and artificial intelligence," IEEE Access, vol. 8, pp. 147728-147737, 2020.

[11] L. Sharma and N. Lohan, "Internet of things with object detection: Challenges, applications, and solutions," in Handbook of Research on Big Data and the IoT: IGI Global, 2019, pp. 89-100.

[12] A. Hazarika, S. Poddar, M. M. Nasralla, and H. Rahaman, "Area and energy efficient shift and accumulator unit for object detection in IoT applications," Alexandria Engineering Journal, vol. 61, no. 1, pp. 795-809, 2022.

[13] W. Cao et al., "CNN-based intelligent safety surveillance in green IoT applications," China Communications, vol. 18, no. 1, pp. 108-119, 2021.

[14] W. Fang, L. Wang, and P. Ren, "Tinier-YOLO: A real-time object detection method for constrained environments," IEEE Access, vol. 8, pp. 1935-1944, 2019.

[15] G. Wang, H. Ding, Z. Yang, B. Li, Y. Wang, and L. Bao, "TRC - YOLO: A real - time detection method for lightweight targets based on mobile devices," IET Computer Vision, vol. 16, no. 2, pp. 126-142, 2022.

[16] A. A. Mei Choo Ang, Kok Weng Ng, Elankovan Sundararajan, Marzieh Mogharrebi, Teck Loon Lim, "Multi-core Frameworks Investigation on A Real-Time Object Tracking Application," Journal of Theoretical & Applied Information Technology, 2014.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779-788.

[18] B. E. Mneymneh, M. Abbas, and H. Khoury, "Vision-based framework for intelligent monitoring of hardhat wearing on construction sites," Journal of Computing in Civil Engineering, vol. 33, no. 2, p. 04018066, 2019.

[19] S. Lu, B. Wang, H. Wang, L. Chen, M. Linjian, and X. Zhang, "A real-time object detection algorithm for video," Computers & Electrical Engineering, vol. 77, pp. 398-408, 2019.

[20] F. Pérez-Hernández, S. Tabik, A. Lamas, R. Olmos, H. Fujita, and F. Herrera, "Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance," Knowledge-Based Systems, vol. 194, p. 105590, 2020.

[21] F. Alrowais et al., "Deep Transfer Learning Enabled Intelligent Object Detection for Crowd Density Analysis on Video Surveillance Systems," Applied Sciences, vol. 12, no. 13, p. 6665, 2022.

[22] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," Multimedia Tools and Applications, pp. 1-33, 2022.

[23] Roboflow. "YOLOv7 Breakdown." https://blog.roboflow.com/pp-yolo-beats-yolov4-object-detection/ (accessed).

[24] C. Liu, Y. Tao, J. Liang, K. Li, and Y. Chen, "Object detection based on YOLO network," in 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), 2018: IEEE, pp. 799-803.

[25] D. Garg, P. Goel, S. Pandya, A. Ganatra, and K. Kotecha, "A deep learning approach for face detection using YOLO," in 2018 IEEE Punecon, 2018: IEEE, pp. 1-4.

[26] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464-7475.

An Energy Efficient Routing Algorithm using Chaotic Grey Wolf with Mobile Sink-based Path Optimization for Wireless Sensor Networks

Latifah ALharthi[1], Alaa E. S. Ahmed[2], Mostafa E. A. Ibrahim[3]
College of Computer and Information Sciences, Imam Muhamad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia[1,2,3]
Shoubra Faculty of Engineering, Benha University, Cairo, Egypt[2]
Department of Electrical Engineering, Benha Faculty of Engineering, Benha University, Benha, Egypt[3]

***Abstract*—The task of deploying an energy-conscious wireless sensor networks (WSNs) is challenging. One of the most effective methods for conserving WSNs energy is clustering. The deployed sensors are divided into groups by the clustering algorithm, and each group's cluster head (CH) is chosen to gather and combine data from other sensors in the group. Mobile Wireless Sensor Networks, which enable moving the sink node, aid in reducing energy consumption. Thus, this paper introduces an energy efficient clustering algorithm and optimized path for a mobile sink using a swarm intelligence algorithms. The Chaotic Grey Wolf Optimization (CGWO) approach is used to form clusters and identify CHs. While utilizing the Slime Mould Algorithm (SMA) for determining the shortest path between a mobile sink and CHs. The effectiveness of the suggested routing strategy is evaluated against that of other current, cutting-edge protocols. The findings demonstrate that in terms of overall energy consumption and network lifetime, the suggested algorithm performs better than others. While for stability period the proposed algorithm outperforms three of compared algorithms and was close to the fourth.**

Keywords—Wireless sensor network; clustering algorithm; grey wolf optimizer; slime mould algorithm; mobile sink

I. Introduction

Mobile Wireless Sensor Networks (MWSNs) enable the movement of entities within a network, functioning as sensor nodes or sinks through mechanisms like wheels, humans, animals, or robots [1, 2]. MWSNs offer a solution to the hotspot problem often encountered in traditional Wireless Sensor Networks (WSNs). In a hotspot scenario, sensor nodes situated near a sink used as a relay tend to deplete their energy rapidly, as the sink increases the communication load on these nearby sensors [1]. MWSNs find applications in various domains, including but not limited to the military, industrial monitoring, habitat observation, healthcare, home networks, disaster management, and security [3]. These applications encompass fire detection systems in forests, battlefield surveillance, traffic monitoring, smart homes and hospitals, pollution control, rescue missions [4] and oil well monitoring.

Clustering algorithms play a crucial role in reducing energy consumption within WSNs. These algorithms partition the sensor nodes into distinct groups or clusters, with each group having a designated cluster head (CH) responsible for coordinating communications between its members and the sink. Clustering can be implemented through various approaches, such as distributed, centralized, or hybrid methods [5]. Sensors consume a significant amount of energy due to their tasks, which include environmental sensing, data transmission, mobility, cluster head (CH) selection, and frequent cluster formation [6]. Additionally, energy demands increase with larger data sizes and greater distances between sensors and the sink.

Numerous algorithms have been proposed to mitigate energy consumption, with clustering being a widely adopted approach. Clustering involves selecting CHs and forming clusters to reduce the number of sensors communicating directly with the sink, thus optimizing communication. Therefore, the process of CH selection is pivotal in clustering. Recent research [6, 7] has explored the use of intelligent swarm algorithms to aid in CH selection, such as ant, firefly, and Grey Wolf Optimization (GWO) algorithms.

This paper investigates the reduction of energy consumption in MWSNs by introducing an enhanced clustering algorithm and optimizing the path for a mobile sink using a swarm intelligence algorithm. Specifically, it employs the Chaotic Grey Wolf Optimization (CGWO) algorithm [8] for CH selection and the Slime Mould Algorithm (SMA) [9] to determine the shortest path between a mobile sink and CHs to reduce energy dissipation, and hence extends the WSN’s life cycle.

The proposed algorithm has the following contributions:

- Employing the CGWO algorithm as a clustering mechanism in MWSNs which to the best of our knowledge has not been investigated up to now in this field.
- Utilizing the SMA algorithm for sink node route determination in MWSNs has not been well studied up to now, and this study aims to fill this gap.
- The results of the proposed algorithm are compared to those of four other state-of-the-art algorithms GWO [10], ACO [11], FA [12], and PSO [13]. in terms of several performance metrics such as network lifetime, stability period, and total consumed energy.

The rest of this paper is structured in five sections. Section II presents the related work. Section III provides the mathematical models for GWO, the enhanced CGWO and Slim Mould Optimization algorithms. Section IV describes the network model and the methodology followed to develop the proposed algorithm. Section V illustrates the simulation results indicating the performance evaluation of the proposed protocol. Finally, Section VI sums up the paper and figure out future directions.

II. Related Work

A. Clustering Algorithms

This section reviews the state-of-the-art clustering algorithms that were recently used in MWSN. The earliest clustering algorithms for WSNs fell into the category of traditional clustering algorithms. These methods employed straightforward techniques for constructing clusters and selecting CHs. In essence, traditional methods designated CHs without the use of sophisticated, intelligent approaches [7]. An example of this approach is the Low Energy Adaptive Clustering Hierarchy (LEACH) [14] protocol.

Authors of [15] presented a heterogeneous clustering algorithm with multiple mobile data collectors (MDCs) to extend the network's lifespan. This algorithm selected CHs using a probability equation based on factors such as energy levels. The MDCs employed the Expectation-Maximization (EM) method to determine optimal paths for CHs based on their positions and energy levels. It demonstrated superior performance, particularly in small areas.

Authors of [11] introduced an enhanced clustering algorithm that employed multiple mobile sinks to improve energy efficiency. This enhanced clustering method incorporated the highest residual energy of sensors as a metric for CH selection, thereby enhancing the traditional LEACH [14] protocol. Mobile sinks utilized the Ant Colony Optimization (ACO) algorithm to identify optimal paths to CHs. Their algorithm defeated the LEACH, Particle Swarm Optimization (PSO), and Genetic Algorithm (GA) in terms of network lifetime and energy consumption.

The authors of [16] proposed inter- and intra-clustering methods to optimize the movements of mobile sinks and conserve energy. The inter-clustering method involved calculating the sojourn time of mobile sinks in clusters, while the intra-clustering method determined the sojourn locations of mobile sinks within clusters. CHs were selected based on proximity to the cluster centre, the highest residual energy was also considered in subsequent CH selections. Their proposed algorithm outperformed the Energy-Efficient PSO-Based Routing algorithm with Mobile Sink (EPMS) and Two-Tier Data Dissemination (TTDD) algorithms in terms of inter-cluster movement.

Recent research papers [7, 17, and 18] have employed optimized clustering algorithms to enhance the network's lifetime. These optimized clustering algorithms leverage Computational Intelligence (CI) methodologies, encompassing fuzzy logic, swarm intelligence, Genetic Algorithms (GA), and petri nets.

In study [19] researchers introduced an algorithm for extending network lifetime and reducing transmission delays. A distributed fuzzy clustering algorithm was used to select CHs. The distributed fuzzy clustering algorithm integrated seven regular components of a fuzzy system into two elements. The first element characterized sensors based on their remaining energy, the number of neighbours, and distances from neighbours. The second element determined sensor and mobile gateway positions, considering factors such as the number of mobile gateways and distances to nearby and distant mobile gateways.

In research [20], authors proposed a fuzzy logic-based algorithm for the clustering process and employed multiple mobile sinks to reduce energy consumption in MWSNs. The fuzzy logic algorithm initially selected temporary CHs and then chose the final CHs from this group based on criteria including distances to the nearest Rendezvous Node (RN), remaining energy, and calculated cluster densities. The sensor areas were divided into regions, each served by a mobile sink. These mobile sinks collected data from the RNs and final CHs using a smart trajectory. Their results showed significant improvement in terms of first node dead, the time at which half of the nodes were still operational (HNA), and the total remaining energy (TRE).

The Particle Swarm Method [21, 22] is a type of swarm intelligence algorithm inspired by the food-searching strategies employed by animal flocks. It divides the swarm into groups, each following a distinct path [21]. Within each group, particles iteratively explore and update information to find the best positions while communicating with others.

In study [23] researchers proposed an enhanced fitness function for the Unequal Clustering PSO (UC-PSO) and Hybrid K-Means Clustering PSO (KC-PSO) algorithms. These improvements aimed to enhance energy efficiency and determine the optimal number of clusters and CHs. The new fitness function selected CHs based on factors such as mobility, residual energy, neighbour connectivity, and distance to the BS. While the KC-PSO algorithm employed UC-PSO in the CH selection process, it utilized a different cluster formation approach. Their results demonstrated that the KC-PSO algorithm outperformed the UC-PSO and LEACH algorithms.

Another example of swarm intelligence algorithms is the Firefly Method [12]. It is used for selecting CHs and introduced a mobile sink to enhance the network's lifetime. CH selection parameters included residual energy, node-to-node distances, and distances from nodes to the sink. The proposed method demonstrated superior performance compared to LEACH, Amend LEACH (A-LEACH), and GA-Based LEACH (LEACH-GA) methods. Additionally, it outperformed the Mobile Sink Improved Energy-Efficient PEGASIS-Based Routing Protocol and Mobile Sink-Based Adaptive Immune Energy-Efficient Clustering Protocol (MSIEEP) algorithms in terms of network lifespan, node residual energy, packet drop ratio (PDR), and packet delays.

The GWO method emulates the social hierarchy observed in grey wolf packs, consisting of alpha, beta, delta, and omega wolves [24]. Researcher in [25] introduced a layered and clustered structure based on the GWO method aimed at optimizing energy consumption.

TABLE I. COMPARISON OF STATE-OF-THE-ART CLUSTERING ALGORITHMS

Ref.	Mobility	Clustering Method/Size [1]	Cluster Head Selection Alg.	Cluster Head Selection Parameters [2]	Simulation Platform	Performance Metrics
[11]	Sink	Dist./dynamic	Traditional	- Residual energy	MATLAB	- Network lifetime - Energy consumption - Average packet loss ratio
[12]	Sink	Cent./dynamic	Firefly	- The residual energy - Distance concerning a node and other nodes - Distance from a node to the sink	MATLAB	- Network lifespan - Residual energy of nodes - Packet drop ratio - Packet delay
[13]	Sink	Cent./dynamic	Traditional	Based on the non-probability Method: - The number of neighbours - The rate at which packets are received	MATLAB	- Number of rendezvous points - Average memory utilization - Number of hops - Packet loss rate - Standard deviation - Throughput - Energy consumption
[15]	Sink	Dist./dynamic	Traditional	- Highest energy	MATLAB	- Number of cluster heads - Network lifetime - Stability period - Throughput
[16]	Sink	Dist./dynamic	Traditional	- The centre of a cluster - Residual energy	OMNet ++ Simulator	- Network lifetime - Residual energy
[19]	Sink	Dist./dynamic	Fuzzy Logic	- The remaining energy - The number of neighbours - The distance to neighbours - The position of a sensor and mobile gateways - The number of mobile gateways - The distance between a sensor and near and distant mobile gateways	OMNet ++ Simulator	- Number of dead sensor nodes - Average remaining energy - Delay in sending packets from the sensor to the base station
[20]	Sink	Dist./dynamic	Fuzzy Logic	- The distance to the closest rendezvous node - The remaining energy - The density	MATLAB	- First node failed - Nodes still operational - Total remaining energy
[22]	Sink	Cent./dynamic	Particle Swarm	- The remaining energy - Centre of the cluster	NA	- Network lifetime - Amount of packet delivery - Energy consumption - Average delivery delay
[23]	All Nodes	Cent./dynamic	Particle Swarm	- Mobility - Residual energy - Neighbours - Distance from the cluster head to the base station	NS2 Simulator	- Number of clusters formed - Network lifetime - Total energy consumption - Packet delivery ratio
[25]	All Nodes	Dist./dynamic	Grey Wolf Optimization	- Residual energy - RSSI - PRR	MATLAB	-Network lifetime - Energy consumption - Throughput
[26]	Sink	Dist./dynamic	Traditional	- Remaining energy - Distance - Data rate	MATLAB	- Network lifetime - Energy consumption - Throughput
[27]	Sink	Dist./dynamic	Traditional	- Residual energy - The ID of a sensor	NS2 Simulator	-Number of active nodes -Average residual energy -Total energy consumption
[28]	Sink	Dist./dynamic	Traditional	- Residual energy - Distance between cluster head and mobile sink	MATLAB	- Network lifetime - Energy consumption
[29]	Sink	Dist./dynamic	Traditional	- The centre of a cluster - Residual energy	Not Mentioned	- Network lifetime - Energy consumption - Packet delivery
[30]	Sink	Cent./dynamic	Traditional	-The centre of a cluster - Residual energy	NS2 Simulator	- Number of alive nodes - Number of delivered packets
[31]	Sink	Dist./dynamic	Traditional	- Residual energy - Distance	MATLAB	- Network lifetime

[1] Cent: Centralized, and Dist: Distributed

[2] Based on weighted probability:

This layered structure consisted of four tiers: alpha, beta, delta, and omega, with the alpha tier being the closest to the static BS. In this context, mobile sensors were analogous to grey wolves, and alpha wolves assumed the role of CHs. CHs were chosen using game theory principles, considering factors such as residual energy, Received Signal Strength Indexes (RSSIs), and Packet Reception Ratios (PRRs). Performance metrics encompassed network lifetime, energy consumption, and throughput. The presented method had better throughput. The results clearly demonstrated that the proposed method outperformed the LEACH protocol. Finally, Table I offers a comparative review of surveyed clustering algorithms indicating the platforms used and other parameters.

B. Wireless Sensor Nodes Mobility

Sink mobility contributes to energy conservation by allowing the sink to move. Path determination is a significant contributor to the energy consumption of WSN. Hence many researchers presented different methods for establishing energy efficient path between nodes [32]. Sink mobility allows for the movement of the sink, and there are three methods for selecting a path between the sink and nodes [2]: random, controlled, and predictable.

First, the easiest way is the random path technique. In the random path method, the mobile sink moves randomly to gather data from sensors [2]. This technique is used in [19, 26]. Second way of moving the sink node is the controlled path method. In the controlled path method, the mobile sink moves strategically within areas that meet certain constraints, such as high residual energy, the number of neighbors, and the number of hops.

Authors of [15] introduced a controlled adaptive mobility model using an EM algorithm. This algorithm assists the sink in collecting data from CHs with the lowest residual energy first. While in study [11] they implemented a controlled path based on the ACO algorithm. Authors of [29] also introduced a controlled route based on improved ACO, considering a distance heuristic factor to enhance its effect on the next node and improve global search ability.

Authors of [16] utilized a controlled trajectory determined by the GA. In [20] they presented a mobile sink that calculates the optimal trajectory by dividing the area of interest into 16 equal parts and considering the average remaining energy of each part. The mobile sink then follows a smart path based on RPs. Researchers of [13] employed the PSO algorithm to determine the path of the mobile sink to CHs. In [33] they presented another technique that integrated ACO and A* algorithms for finding the best energy efficient route between CHs and a base station.

Third way of sink movement is the predictable path. In the predictable path method, the mobile sink follows a predefined route to specific relay or data collector nodes responsible for gathering data from sensor nodes and transmitting it to the mobile sink [2]. Authors of [12] introduced a mobile sink that selects its path by dividing the network into four or eight areas and moves to each part using the centroid of the CHs. While [27] used a predictable trajectory. Authors of [28] used a predetermined route depending on the angular velocity.

III. Technical background

This section provides the mathematical models for Grey Wolf Optimization (GWO), the enhanced Chaotic GWO and Slim Mould Optimization algorithms.

A. Original GWO

The original GWO algorithm, introduced by [34], draws inspiration from the cooperative hunting and social hierarchy of grey wolves to tackle optimization problems. The core concept of the GWO algorithm involves locating a target, or “prey”, by mimicking the leadership hierarchy of grey wolves. The inspiration is in the cooperative hunting and social hierarchy of grey wolves.

Grey wolves organize themselves into a dominant social hierarchy, featuring alpha, beta, delta, and omega wolves. The alpha wolves occupy the top tier of this hierarchy, where they make decisions regarding hunting and habitat selection. The beta wolves comprise the second tier and have the authority to issue commands to the delta and omega wolves. Delta wolves, in turn, follow the directives of the alpha and beta wolves and oversee the omega wolves. Ultimately, the omega wolves obediently follow the commands of all other members of the pack.

The population-based meta-heuristic method known as "grey wolf optimization" (GWO) mimics the natural hunting strategy and leadership structure of grey wolves. The GWO hunting process consists of several key phases:

- Tracking, chasing, and approaching the prey.
- Pursuing, encircling, and harassing the prey until it stops moving.
- Attacking the prey.

1) GWO mathematical model: The GWO algorithm [24, 34] considers the fittest solution as the alpha (α). As a result, the second and third-best solutions are designated as beta (β) and delta (δ), respectively. The remaining candidate solutions are assumed to be omega (ω). α, β, and δ guide the hunting process, with the ω wolves following these three leaders.

a) Encircling prey: Mathematically, grey wolves enclose and surround their prey as follows [34]:

$$\vec{\mathrm{D}} = \left|\vec{\mathrm{C}} \cdot \vec{\mathrm{X}}_p(\mathrm{t}) - \vec{\mathrm{X}}(\mathrm{t})\right| \tag{1}$$

$$\vec{\mathrm{X}}(\mathrm{t}+1) = \vec{\mathrm{X}}_p(\mathrm{t}) - \vec{\mathrm{A}} \cdot \vec{\mathrm{D}} \tag{2}$$

where, $\vec{\mathrm{D}}$ indicates the distance between prey and wolf, t corresponds to the present iteration, $\vec{\mathrm{X}}(\mathrm{t})$ is the current position of the wolf, and $\vec{\mathrm{X}}_p(\mathrm{t})$ is the position of the prey.$\vec{\mathrm{A}}$ and $\vec{\mathrm{C}}$ are coefficient-vectors, $\vec{\mathrm{X}}$p is the vector’s location of the prey, and X is the vector’s location of a grey wolf. vectors $\vec{\mathrm{A}}$ and $\vec{\mathrm{C}}$ can be computed as follows:

$$\vec{\mathrm{A}} = 2\,\vec{\mathrm{a}} \cdot \vec{\mathrm{r}}_1 - \vec{\mathrm{a}}, \qquad \vec{\mathrm{C}} = 2 \cdot \vec{\mathrm{r}}_2 \tag{3}$$

where, $\vec{\mathrm{r}}_1$ and $\vec{\mathrm{r}}_2$ are random vectors in range of [0,1], and the components of $\vec{\mathrm{a}}$ are reduced linearly from 2 down to 0 throughout repeated iterations.

b) Hunting: The algorithm keeps the first three best solutions alpha α, beta β, and delta δ and obliges the omega ω wolves to adjust their positions based on positions of wolves α, β, and δ. Eq. (4) to Eq. (6) indicates how the distances from α, β, and δ wolves ($\vec{D}_\alpha$, $\vec{D}_\beta$ and $\vec{D}_\delta$) to each of the lasting wolfs, using positions α, β, and δ wolves ($\vec{X}_\alpha$, $\vec{X}_\beta$ and $\vec{X}_\delta$) and the position of lasting wolves($\vec{X}$):

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right|, \vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right| \qquad (4)$$

$$\vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}\right|$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \left(\vec{D}_\alpha\right), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \left(\vec{D}_\beta\right), \qquad (5)$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \left(\vec{D}_\delta\right)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \qquad (6)$$

c) Attacking prey: In study [34] Grey wolves conclude the hunt by attacking the prey when it stops moving. The GWO algorithm models approaching the prey mathematically by decreasing the value of $\vec{a}$ and narrowing the fluctuation range of $\vec{A}$ where |A|<1 force the wolves to attack the prey (exploitation).

d) Deletion: The search is based on the positions of the alpha, beta, and delta grey wolves. Grey wolves diverge from each other to search for prey and converge to attack prey. In the mathematical model of divergence, the GWO algorithm employs $\vec{A}$ where |A| > 1 to compel the search agents to diverge from the prey. The $\vec{C}$ vector contains random values in [0, 2], providing random weights for prey. This stochastic weighting either emphasizes (when C > 1) or deemphasizes (when C < 1) the attack. It reflects the effect of obstacles to approaching prey in nature. Depending on the position of a wolf, it can randomly assign weight to the prey, making it harder or easier for wolves to reach it.

B. Chaotic GWO

CGWO algorithm utilizes a varying number of wolves, which are regarded as leaders during each iteration. Rodrigues (2021) [8] proposed the use of a chaotic variable to determine the number of leaders in the pack during each iteration. To calculate the number of leader wolves for each iteration, the following formula is employed:

$$n(t) = \left\lceil \frac{M.z(t)}{2} \right\rceil \qquad (7)$$

The number of leader wolves n(t) in each iteration ranges between 1 and half of the population size. The author used the Celling function for n(t) to round the result to the next integer. Here, M represents the number of wolves in the pack, z(t) is a chaotic variable within the interval [0,1] [8, 35]. In contrast to the way of updating the position of individual wolves in GWO described by Eq. (6), in CGWO the author utilized Eq. (8).

$$X(t+1) = \frac{\sum_{v=1}^{n(t)} X_v}{n(t)}, X_v = X_j(t) - A_j . D_j \qquad (8)$$

where, Xj(t) represents the position of the wolf with the j-th best fitness value in iteration t, Aj is a random vector computed according to Eq. (4), (3) and D_j is calculated using the following equation:

$$D_j = \left|C_j . X_j(t) - X\right| \qquad (9)$$

where, X is the current position of the wolf, and C_j is a random vector calculated according to Eq. (3).

Chaotic maps [8] represent the chaotic function, often referred to as an orbit. An orbit is an iterative function that generates a sequence of values in each iteration. The characteristics of an orbit are its aperiodic nature, boundedness (chaotic variables have upper and lower limits), and sensitivity to initial conditions.

The CGWO algorithm, in contrast to GWO, incorporates a chaotic sequence to determine a varying number of leaders in each iteration. This approach enhances the CGWO algorithm's diversification capability and strikes a balance between diversification and intensification capabilities, which is crucial for the optimization algorithm's overall performance. By involving a larger number of wolves as leaders, candidate solutions that are not near the optimal solution contribute to guiding the search process. In [8] nine chaotic maps are investigated.

C. Slim Mould Optimization Algorithm

The slime mould organism relies on creating an interconnected venous network to seek out food sources, allowing it to generate optimal paths for reaching food [9]. The SMA algorithm has two phases:

1) Approach food: The organic matter in slime mould seeks food, surrounds it, and secretes enzymes to digest it. The slime mould navigates towards a food source using a mathematical expression designed to mimic its contraction behavior as follows:

$$\overrightarrow{X(t+1)} = \begin{cases} \overrightarrow{X_b(t)} + \overrightarrow{vb} \cdot \left(\vec{W} \cdot \overrightarrow{X_A(t)} - \overrightarrow{X_B(t)}\right), r < p \\ \vec{vc} \cdot \overrightarrow{X(t)}, r \geq p \end{cases} \qquad (10)$$

where, $\overrightarrow{vb}$ is a parameter with a range of $[-a, a]$, $\vec{vc}$ decreases linearly from 1 to 0. t indicates the current iteration, $\overrightarrow{X_b}$ denotes the individual location with the highest odour concentration currently found, $\vec{X}$ indicates the location of the slime mould, $\overrightarrow{X_A}$ and $\overrightarrow{X_B}$ indicate two individuals randomly selected from the swarm, and $\vec{W}$ indicates the weight of the slime mould. P is computed as follow:

$$p = \tanh|S(i) - DF| \qquad (11)$$

where, $i \in 1,2, \dots, n$, S(i) denotes the fitness of $\vec{X}$ and DF indicates the best fitness obtained in all iterations. $\overrightarrow{vb}$ is computed as:

$$\overrightarrow{vb} = [-a, a], a = \operatorname{arctanh}(-\left(\frac{t}{max\_t}\right) + 1) \qquad (12)$$

where, max_t indicates the maximum number of iterations. $\vec{W}$ is calculated as:

$$\overrightarrow{W(SI(\imath))} = \begin{cases} 1 + r \cdot \log\left(\frac{bF - S(i)}{bF - wF} + 1\right), condition \\ 1 - r \cdot \log\left(\frac{bF - S(i)}{bF - wF} + 1\right), \text{ others,} \end{cases} , \qquad (13)$$

$$SI = sort(S)$$

Where condition denotes that S(i) ranks in the first half of the population, r represents a random value within the interval [0,1], bF stands for the optimal fitness obtained in the current iterative process, wF indicates the worst fitness value obtained in the current iterative process, and SI represents the sequence of fitness values sorted in ascending order.

2) Wrap food: It mimics the contraction mode of the venous tissue structure of slime mould during its search for food. As the concentration of food encountered by the vein increases, the bio-oscillator generates stronger waves, resulting in faster cytoplasmic flow and thickening of the vein [9]. The slime mould updates its location using the following mathematical formula:

$$\overrightarrow{X^*} = \begin{cases} \mathrm{rand} \cdot (\mathrm{UB} - \mathrm{LB}) + \mathrm{LB}, \mathrm{rand} < z \\ \overrightarrow{X_b(t)} + \overrightarrow{vb} \cdot \left(W \cdot \overrightarrow{X_A(t)} - \overrightarrow{X_B(t)}\right), r < p \\ \overrightarrow{vc} \cdot \overrightarrow{X(t)}, r \geq p \end{cases} \tag{14}$$

where, LB and UB represent the lower and upper boundaries of the search range, while rand and r denote random values in the range [0,1]. The constant z is set to 0.03.

3) Osillation: Slime mould relies on a biological oscillator to generate propagating waves that alter the flow of cytoplasm within its veins, allowing it to position itself more effectively in areas of higher food concentration. The vector value $\overrightarrow{vb}$ randomly oscillates within the range of $[-a, a]$ and gradually approaches 0 with increasing iterations. Similarly, the vector value $\overrightarrow{vc}$ oscillates within the range of [–1,1] and tends to 0 eventually [9]. Using the SMA algorithm, the mobile sink chooses the shortest path between itself and the CHs to collect sensed data.

IV. Methodology

This section dives into the design and implementation of the proposed algorithm. The proposed algorithm consists of two phases: cluster construction and path formation for a mobile sink. It combines the CGWO algorithm for the selection of CHs and the SMA for determining the shortest route between a mobile sink and the CHs.

A. WSN Model

The wireless sensor network model considered in this paper has the following assumptions:

- The network model is synchronous.
- All nodes are stationary but the mobile sink station.
- All nodes have the same initial battery capacity and can perform the same functions.
- Each sensor node is identified with a unique identifier.
- The links are symmetric.
- The distance between sensor nodes can be calculated based on received signal strength indicator (RSSI).

B. Cluster Structure Phase

The main functions of this phase are electing CHs using the CGWO algorithm and associating sensor nodes with respective CHs to form clusters.

The sensor nodes are randomly deployed in a region, and a mobile sink is located in the center of the network. Initially, the sensor nodes send their locations and remaining energy to a mobile sink. Then, the mobile sink selects the CHs based on the fitness function of the CGWO algorithm. The clusters have the following properties:

- Centralized clustering formation method: The mobile sink utilizes the CGWO algorithm to select CHs.
- Fixed cluster count: It specifies a fixed number of clusters, which is ten clusters in each round.
- Variable cluster size: The number of cluster members is not fixed in each round.
- Intra-cluster topology: The proposed algorithm relies on single hops to connect cluster members to their respective CHs.
- Inter-CH connectivity: The proposed algorithm establishes direct connections from CHs to the mobile sink.

This article introduced a new fitness function that relies on the following parameters for selecting CHs:

- Remaining energy of the CH.
- The CH’s membership count.
- Euclidean distance from a mobile sink to a CH.
- CH centrality.

The fitness function for sensor node i is calculated as follows:

$$F(i) = aF_f(i) + (1 - a)F_u(i) \tag{15}$$

$$F_f(i) = R_e(i) + S_{m(i\,)} \tag{16}$$

$$F_u(i) = [E_u(i, M_s)]^{-1} + C_s(i) \tag{17}$$

where, a is a scaling factor with a value from 0.1 to 0.9, F_f is the fundamental fitness function, and F_u represents the non-fundamental fitness function.

The fundamental fitness function (F_f) calculates the sensor node’s members (S_m) and its remaining energy (R_e). Sm signifies the number of connecting nodes to a particular node within its transmission range, while R_e is the ratio of the remaining energy to the initial energy of the node.

The non-fundamental fitness function (F_u) computes the Euclidean distance (E_u) and sensor centrality (C_s). E_u calculates the Euclidean distance from sensor i to the mobile sink M_s, while C_s determines the sensor node’s centrality among its neighbors.

The sensor’s members (Sm) is computed as follows:

$$N(i) = \sum_{j \in W} b_{ij} \tag{18}$$

where, W is the wireless sensor network, $b_{ij} = 1$ means that distributed sensor i is connected to the distributed sensor j; otherwise, $b_{ij} = 0$.

Algorithm 1: Cluster Formation Algorithm

```
A.  Initialize WSN parameters: Area, number of nodes, initial mobile sink
    position (50,50), … etc.
Input:
    Number of alive nodes
    Number of packs $P_i$
    Number of cluster heads in a pack: 10% of the total number of alive
    sensor nodes randomly.
Output:
    Cluster heads
B.  Cluster head selection using CGWO algorithm.
    Initialize the population $X_i$, choose 10% of sensor nodes as cluster
    heads in a pack from alive nodes randomly.
    Initialize a = 2.
    Initialize vectors A and C.
    Initialize the Tent chaotic map.
    Compute the fitness value of each wolf according to Eq. (15).
    Define the number of leaders n(t) according to Eq. (7).
    For j = 1 : n(t) do
        Compute $D_j$ according to Eq. (9).
        Compute $X_v$ according to Eq. (8).
    End for.
    While t < maximum number of iterations do
        For each wolf i do
            Update its position X(t+1) according to Eq. (8).
        End for.
        Update a, A, C, and n(t).
        Compute the new fitness value of each wolf.
        For j = 1 : n(t) do
            Compute $D_j$ according to Eq. (9).
            Compute $X_v$ according to Eq. (8).
        End for
        Increment the iteration number (t = t + 1).
    End while
    Return the best solution (i.e. cluster head)
    CH = Set of sensor nodes in $P_i$.
End Algorithm.
```

The sensor's members (C_s) is computed as follows:

$$C_s(i) = \frac{(N-1)}{\sum_{j=1}^{N} d_{ij}} \tag{19}$$

where, N is the number of sensors and d represents the shortest distance from sensor i to sensor j.

The remaining energy of sensor node (R_e) is computed as follows:

$$\mathbf{R_e} = \frac{\mathbf{1}}{\sum_{j=1}^{m}\left(E_{CH_j}\right)} \tag{20}$$

where, m is the total number of CHs, l_j is the number of sensor nodes in cluster j, and ECH_j is the current energy of CH_j, $1 \leq j \leq m$.

The Euclidean distance (E_u) is computed as follows is computed as follows:

$$E_u = \sum_{j=1}^{m}\left(\frac{1}{l_j}\mathrm{dis}\left(CH_j, M_s\right)\right) \tag{21}$$

where, $\mathrm{dis}(CH_j, M_s)$ signifies the distance between CH_j and M_s.

Based on the experimental investigations, this paper employs the tent function Eq. (22) for the chaotic map function and set 'a' to 0.2 in the fitness function to achieve superior outcomes.

$$z(t+1) = \begin{cases} z(t)/0.4, & 0 < z(t) \leq 0.4 \\ \left(1 - z(t)\right)/0.6, & 0.4 < z(t) \leq 1 \end{cases} \tag{22}$$

In this study, the sensor nodes will transmit their locations and remaining energy to the mobile sink. Subsequently, the mobile sink will select CHs using the CGWO algorithm as described in Algorithm 1. Following this selection, the sensor nodes will align themselves with their respective CHs and begin transmitting the sensed data.

C. Path Formation Phase

Following the election of CHs, the mobile sink employs the SMA algorithm to move towards the nearest CH, followed by the second closest CH, and so forth, for data collection. This algorithm exerts control over the movement of the mobile sink.

Algorithm 2: Mobile Sink Path Formation Algorithm

```
Input: Number of cluster heads which is determined by the Algorithm
1

Output: The slime mould position = the mobile sink position

Initialize the population size = number of cluster heads.
Initialize the positions of the slime mould positions.
While t ≤ Max_iteration do
    Calculate the fitness of all the slime moulds by Eq. (23).
    Update bestFitness, $x_b$.
    Calculate the W by Eq. (13).
    For each search portion do

        Update p, vp, vc.
        Update positions by Eq. (14).
    End For
t = t + 1.
End while
Return best Fitness, $x_b$ = the mobile sink position.
```

A new fitness function is proposed for determining the path between a mobile sink and CHs. The following equation calculates the path distance (D) travelled by the mobile sink (M_S):

$$D(T_i) = \mathrm{dis}(M_S, ch_1) + \sum_{i=1}^{e-1}\mathrm{dis}(ch_i, ch_{i+1}) + \mathrm{dis}(ch_e, M_S) \tag{23}$$

where, ch is the number of CHs, ch1 is the nearest CH to the mobile sink MS, ch_e is the farthest CH, and $\mathrm{dis}(M_S, ch_1)$ indicates the distance between CHs or between a CH and an M_S. The M_S begins its journey from an initial position, visits all CHs, and returns to the starting point.

Algorithm 2 outlines the steps involved in constructing the mobile sink's route using the SMA algorithm.

D. Energy Consumption Model

This paper utilizes the first-order radio model as the energy consumption model for both sending and receiving data, as proposed by [36]. This model incorporates both the free space (f_s) and multi-path fading (m_p) models and is contingent on the distance between the sender and receiver. When the distance is less than the threshold value d_0, the authors employ the free space (f_s) model. Otherwise, the authors switch to the multi-path (m_p) model. The energy consumption for transmitting an l-bit message over a distance d is calculated as follows:

$$E_{Tx}(l,d) = \begin{cases} lE_{elec} + lE_{fs}\, d^2, & d < d_0 \\ lE_{elec} + lE_{mp}\, d^4, & d \geq d_0 \end{cases} \qquad (24)$$

where, E_{Tx} represents the total energy required for transmission, E_{elec} denotes the energy dissipation per bit for circuit operation, including the transmitter or receiver, E_{fs} is the energy used for amplification in the free space model, and Emp pertains to the multi-path model and is significantly influenced by the transmitter amplifier model. The energy required to receive an l-bit message is calculated as:

$$E_{Rx}(l) = E_{Rx-elec}(l) = lE_{elec} \qquad (25)$$

where, E_{Rx} represents the energy consumption for data reception. The threshold distance d_0 is set as follows:

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \qquad (26)$$

V. Simulation Experments and Results

A. Experimintal Results

All the simulation experiments were conducted on a laptop with a processor speed of 1.70 GHz Intel Core i7 and memory of 16 GB. The operating system used was Windows 11, version 22H2. The simulation platform was the MATLAB 2023a. A WSN of area 100× 100 m^2 is considered. The nodes are uniformly deployed with the BS initially in the center of the network area. Table II presents the network model and simulation parameters used in the experiments.

TABLE II. Network Setup and Simulation Parameters

Parameter	Value
Simulation area	100 × 100 m^2
Total number of nodes	100
Initial energy	0.5 Joules
Number of rounds	2500
Packet size	2000 bytes
Initial position of the mobile sink	(50,50)
E_{elec}	$5.0e^{-8}$
E_{fs}	$1.0e^{-11}$
E_{mp}	$1.3e^{-15}$

B. Simulation Results

This section describes and compares the developed algorithm's performance with that of the following: FA [12], GWO [10], ACO [11], and PSO [13]. Several measures, including total residual energy, total energy consumption, network lifetime, and stability period are used for analysing and assessing the proposed algorithm.

The total residual energy versus time (in terms of rounds) is presented in Fig. 1. The suggested protocol exhibits larger residual energy than the other three algorithms, as the simulation results obviously reveal. Notably, the proposed algorithm outperforms compared ones, with the residual energy reaching 0 at rounds 1142, 2318, 2381, and 2381 for ACO [11], FA [12], PSO [13], and GWO [10], respectively. On the other hand, the proposed algorithm reaches 0 at round 2490.

The developed algorithm maintains the highest energy levels until 2489 rounds due to its use of the optimized clustering chaotic variable in conjunction with the GWO algorithm for CH selection. Additionally, the SMA algorithm aids the CGWO algorithm by facilitating the mobile sink's path determination to CHs.

Fig. 2 represents the total energy consumption versus time (in terms of rounds) of the proposed algorithm compared to the three other algorithms. The total energy consumption is measured during both transmission and reception and is calculated by dividing total energy consumption by total initial energy. Remarkably, the developed algorithm demonstrates significantly lower energy consumption compared to compare algorithms.

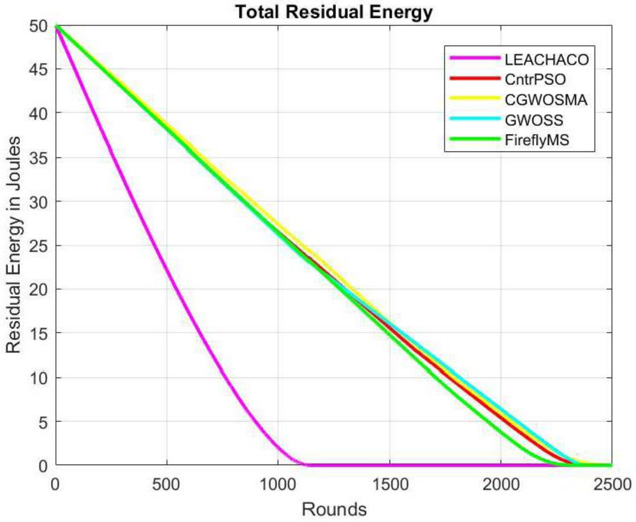


Fig. 1. Total residual energy versus time.

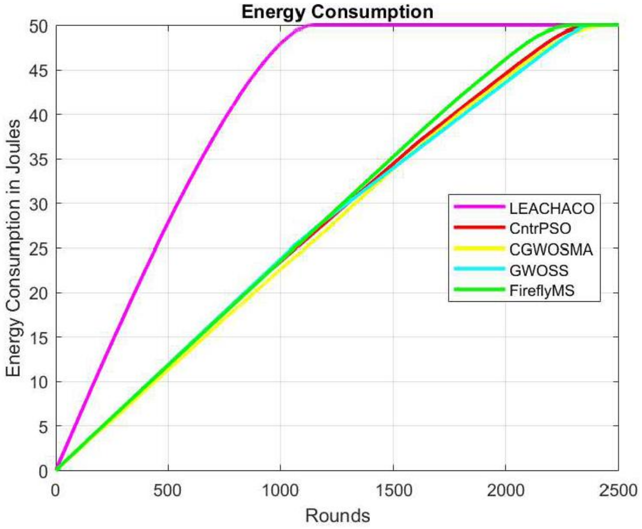


Fig. 2. Total energy consumption versus time.

On average, the energy consumption when using the proposed algorithm is lower than that of ACO [11], FA [12], PSO [13], and GWO [10] by 57.71%, 4.30%, 3.07%, and 2.46%, respectively.

Network lifetime refers to the duration for which nodes remain operational. It signifies the time span from the initiation of network operation until its conclusion, which is marked by the depletion of the last functioning sensor node [37].

Fig. 3 illustrates that it is evident that the developed algorithm sustains more alive nodes than other compared algorithms. The network lifetime is represented in terms of the percentage of alive nodes. ACO [11], PSO [13], GWO [10] algorithms, the proposed algorithm, and FA [12] algorithm, maintain 100% of alive nodes until rounds 159, 836, 1061, 1525, and 1619, respectively. However, ACO [11], FA [12], PSO [13] algorithms, the developed algorithm, and GWO [10] algorithm reach 50% of alive nodes at rounds 761, 2179, 2269, 2294, and 2343, respectively. Finally, ACO [11], FA [12], PSO [13], GWO [10] algorithms, and the implemented algorithm dwindle to 0% of alive nodes at rounds 1143, 2318, 2381, 2381, and 2490, respectively.

Thus, on average, when utilizing our proposed algorithm, the network lifetime surpasses that of ACO [11], FA [12], PSO [13], and GWO [10] by 318.17%, 34.31%, 6.85%, and 1.44%, respectively. The results highlight that employing the CGWO algorithm in the clustering formation significantly enhances network lifetime. The selection of CHs is influenced by chaotic variable, thereby extending the network's operational duration.

The primary driver behind the developed algorithm's superior performance is that the developed algorithm employs the CGWO algorithm, utilizing chaos variable. The developed algorithm utilizes a fitness function that considers both the residual energy and centrality of CHs when choosing them. While the SMA algorithm employs a fitness function based on the shortest distance between the mobile sink and CHs, contributing to enhanced energy retention. While the GWO algorithm [10] exhibits lower residual energy compared to the developed algorithm but fares better than the other compared algorithms. This is because it shares similarities with the developed algorithm in CH selection but uses the original GWO algorithm.

The PSO algorithm in [13] maintains higher residual energy than the FA and ACO algorithms, as it employs the PSO algorithm to determine the shortest path between the mobile sink and CHs. However, it lags behind the GWO and the developed algorithm because it does not use the optimized clustering approach for CH selection; it merely chooses nodes based on centrality. The FA algorithm [12] exhibits higher residual energy than the ACO algorithm as it selects CHs using the FA algorithm and employs a predictable path for the mobile sink. Nonetheless, it falls short of the PSO and GWO algorithms as well as the developed algorithm. Finally, the ACO algorithm [11] ranks the lowest due to its use of the traditional clustering approach LEACH. Although it utilizes three mobile sinks that employ the ACO algorithm to determine paths between mobile sinks and CHs, it lags significantly behind in energy retention compared to the other algorithms.

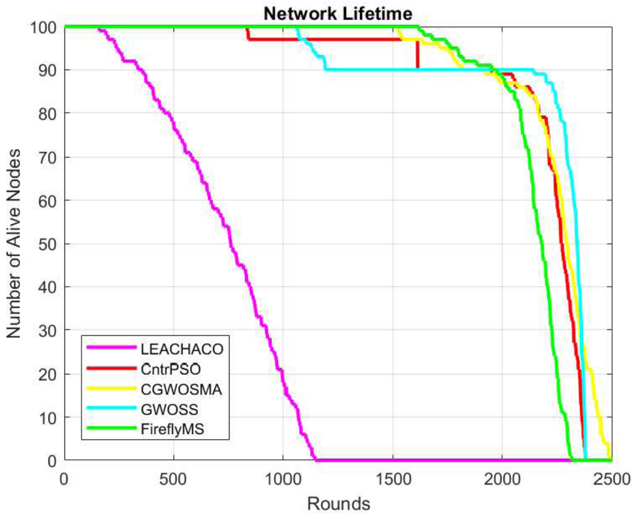


Fig. 3. Number of alive nodes versus time.

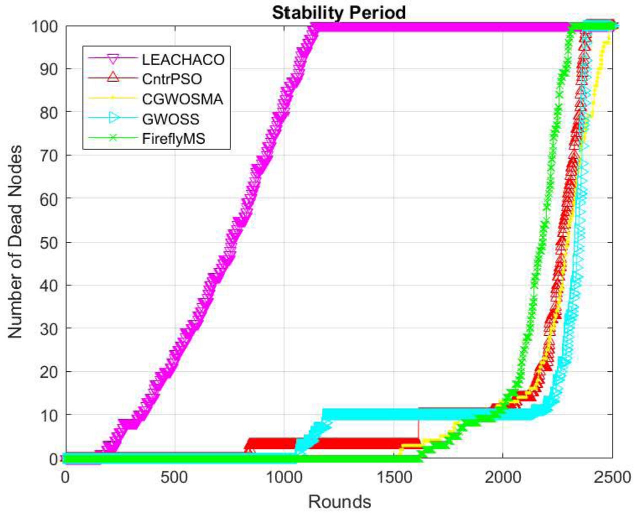


Fig. 4. Number of died nodes versus time.

At last, the stability period for all algorithms is measured. The stability period is defined as "The time when the first node died" [38]. Fig. 4 depicts when the first node in each algorithm died. Specifically, in the ACO [11], PSO [13], GWO [10] algorithms, the implemented algorithm, and FA [12] algorithm, the first node's energy depletion occurs at rounds 160, 837, 1062, 1526, and 1620, respectively.

Fig. 4 reveals that the FA [12] algorithm exhibits greater stability compared to the other algorithms. This enhanced stability can be attributed to its strategy of selecting CHs based on the FA algorithm and utilizing a predictable path for the mobile sink.

On the other hand, the implemented algorithm is less stable than the FA algorithm but more stable than the other two compared algorithms. This is primarily because it depends on node centrality in the fitness function, which is used in CH selection. This can lead to the early depletion of energy in nodes farthest from the CHs. Additionally, the fitness function of the SMA algorithm is based solely on the shortest distance from the mobile sink to the CHs.

VI. Conclusion and Future work

Clustering is one of the best techniques for WSN energy conservation. The clustering method divides the deployed sensors into groups, and each group's cluster head (CH) is selected to collect and aggregate data from other group members. Energy consumption can be decreased with the use of mobile wireless sensor networks, which allow the sink node to be moved. As a result, this paper suggests using swarm intelligence to cluster wireless sensor networks and select dynamic routes for mobile sinks.

To create clusters and locate CHs, the Chaotic Grey Wolf Optimization (CGWO) technique is employed. When figuring out the shortest route between a mobile sink and CHs using the Slime Mould Algorithm (SMA). This paper introduces a new fitness function that relies on remaining energy of the CH, the CH's membership count, Euclidean distance from a mobile sink to a CH, and CH centrality for selecting CHs and for determining the path between a mobile sink and CHs.

The performance of the proposed technique is compared with various state-of-the-art protocols. The results show that the recommended algorithm outperforms other compared algorithms in terms of overall energy consumption and network longevity. The developed approach performs better than three of the compared algorithms and is nearly as good as the fourth throughout the stability period.

As a future work, the limitations of the developed algorithm including the following aspects will be investigated. First, to improve stability period performance by refining the fitness function within the CGWO algorithm and exploring alternative swarm intelligence methods for both clustering and path formation, second to intend to employ the implemented algorithms with other mobility models where more than sink node is mobile.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) for funding this work through Graduate Students Research Support Program.

References

[1] Obaidat, M., & Misra, S. (2014). Wireless mobile sensor networks. In Principles of Wireless Sensor Networks (pp. 248-281). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139030960.012

[2] Sara, G. S., & Sridharan, D. (2014). Routing in mobile wireless sensor network: A survey. Telecommunication Systems, 57, 51-79. Available at: https://doi.org/10.1007/s11235-013-9766-2.

[3] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). A survey on sensor networks. IEEE Communications magazine, 40(8), 102-114. Available at: https://doi.org/10.1109/MCOM.2002.1024422.

[4] Tunca, C., Isik, S., Donmez, M. Y., & Ersoy, C. (2013). Distributed mobile sink routing for wireless sensor networks: A survey. IEEE communications surveys & tutorials, 16(2), 877-897. Available at: https://doi.org/10.1109/SURV.2013.100113.00293.

[5] Jain, N., Sinha, P., & Gupta, S. K. (2013). Clustering protocols in wireless sensor networks: A survey. International Journal of Applied Information System (IJAIS), 5(2).

[6] Sumathi, J., & Velusamy, R. L. (2021). A review on distributed cluster based routing approaches in mobile wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing, 12, 835-849. Available at: https://doi.org/10.1007/s12652-020-02088-7.

[7] Wohwe Sambo, D., Yenke, B. O., Förster, A., & Dayang, P. (2019). Optimized clustering algorithms for large wireless sensor networks: A review. Sensors, 19(2), 322. Available at: https://doi.org/10.3390/s19020322.

[8] Rodrigues, L. R. (2023). A chaotic grey wolf optimizer for constrained optimization problems. Expert Systems, 40(4), e12719. Available at: https://doi.org/10.1111/exsy.12719.

[9] Li, S., Chen, H., Wang, M., Heidari, A. A., & Mirjalili, S. (2020). Slime mould algorithm: A new method for stochastic optimization. Future Generation Computer Systems, 111, 300-323. Available at: https://doi.org/10.1016/j.future.2020.03.055.

[10] Agrawal, D., Wasim Qureshi, M. H., Pincha, P., Srivastava, P., Agarwal, S., Tiwari, V., & Pandey, S. (2020). GWO-C: Grey wolf optimizer-based clustering scheme for WSNs. International Journal of Communication Systems, 33(8), e4344. Available at: https://doi.org/10.1002/dac.4344.

[11] Krishnan, M., Yun, S., & Jung, Y. M. (2019). Enhanced clustering and ACO-based multiple mobile sinks for efficiency improvement of wireless sensor networks. Computer Networks, 160, 33-40. Available at: https://doi.org/10.1016/j.comnet.2019.05.019.

[12] Chauhan, V., & Soni, S. (2020). Mobile sink-based energy efficient cluster head selection strategy for wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing, 11, 4453-4466. Available at: https://doi.org/10.1007/s12652-019-01509-6.

[13] Tabibi, S., & Ghaffari, A. (2019). Energy-efficient routing mechanism for mobile sink in wireless sensor networks using particle swarm optimization algorithm. Wireless Personal Communications, 104, 199-216. Available at: https://doi.org/10.1007/s11277-018-6015-8.

[14] Heinzelman, W. R., Chandrakasan, A., & Balakrishnan, H. (2000, January). Energy-efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd annual Hawaii international conference on system sciences (pp. 10-pp). IEEE.

[15] Toor, A. S., & Jain, A. K. (2019). Energy aware cluster based multi-hop energy efficient routing protocol using multiple mobile nodes (MEACBM) in wireless sensor networks. AEU-International Journal of Electronics and Communications, 102, 41-53. Available at: https://doi.org/10.1016/j.aeue.2019.02.006.

[16] Gharaei, N., Bakar, K. A., Hashim, S. Z. M., & Pourasl, A. H. (2019). Inter-and intra-cluster movement of mobile sink algorithms for cluster-based networks to enhance the network lifetime. Ad Hoc Networks, 85, 60-70. Available at: https://doi.org/10.1016/j.adhoc.2018.10.020.

[17] Ahmed, A. E., & Ibrahim, M. E. (2018). Colored Petri Net Models for Clustered and Tree-Based Data Aggregation in Wireless Sensor Networks. International Journal on Information Technologies & Security, 10(3).

[18] Engelbrecht, A. P. (2007). Computational intelligence: an introduction. John Wiley & Sons.

[19] Abdolkarimi, M., Adabi, S., & Sharifi, A. (2018). A new multi-objective distributed fuzzy clustering algorithm for wireless sensor networks with mobile gateways. AEU-International Journal of Electronics and Communications, 89, 92-104. Available at: https://doi.org/10.1016/j.aeue.2018.03.020.

[20] Koosheshi, K., & Ebadi, S. (2019). Optimization energy consumption with multiple mobile sinks using fuzzy logic in wireless sensor networks. Wireless Networks, 25, 1215-1234. Available at: https://doi.org/10.1007/s11276-018-1715-2.

[21] Al Aghbari, Z., Khedr, A. M., Osamy, W., Arif, I., & Agrawal, D. P. (2020). Routing in wireless sensor networks using optimization techniques: A survey. Wireless Personal Communications, 111, 2407-2434. Available at: https://doi.org/10.1007/s11277-019-06993-9.

[22] Wang, J., Cao, Y., Li, B., Kim, H. J., & Lee, S. (2017). Particle swarm optimization based clustering algorithm with mobile sink for WSNs. Future Generation Computer Systems, 76, 452-457. Available at: https://doi.org/10.1016/j.future.2016.08.004.

[23] Sangeetha, M., & Sabari, A. (2018). Prolonging network lifetime and optimizing energy consumption using swarm optimization in mobile

wireless sensor networks. Sensor Review, 38(4), 534-541. Available at: https://doi.org/10.1108/SR-08-2017-0157.

[24] Cuevas, E., Fausto, F., González, A., Cuevas, E., Fausto, F., & González, A. (2020). An introduction to nature-inspired metaheuristics and swarm methods. New Advancements in Swarm Algorithms: Operators and Applications, 1-41. Available at: https://doi.org/10.1007/978-3-030-16339-6_1.

[25] Raajini, X. M., Kumar, R. R., Indumathi, P., & Praveen, V. Grey-Wolf Optimization Approach for Routing in Dynamic Wireless Sensor Network. International Journal of Applied Engineering Research, 10(32), 2015.

[26] Saranya, V., Shankar, S., & Kanagachidambaresan, G. R. (2019). Energy efficient data collection algorithm for mobile wireless sensor network. Wireless Personal Communications, 105, 219-232. Available at: https://doi.org/10.1007/s11277-018-6109-3.

[27] Amini, S. M., Karimi, A., & Shehnepoor, S. R. (2019). Improving lifetime of wireless sensor network based on sinks mobility and clustering routing. Wireless Personal Communications, 109, 2011-2024.

[28] Wang, J., Gao, Y., Liu, W., Sangaiah, A. K., & Kim, H. J. (2019). Energy efficient routing algorithm with mobile sink support for wireless sensor networks. Sensors, 19(7), 1494. Available at: https://doi.org/10.3390/s19071494.

[29] Wang, J., Cao, J., Sherratt, R. S., & Park, J. H. (2018). An improved ant colony optimization-based approach with mobile sink for wireless sensor networks. The Journal of Supercomputing, 74(12), 6633-6645. Available at: https://doi.org/10.1007/s11227-017-2115-6.

[30] Mechta, D., Harous, S., & Alem, I. (2018). Improving wireless sensor networks durability through efficient sink motion strategy: TMSRP. Wireless Personal Communications, 99, 1661-1682. Available at: https://doi.org/10.1007/s11277-018-5329-x.

[31] Pandey, S., & Anand, V. (2017). Load-balanced clustering scheme with sink mobility for heterogeneous wireless sensor networks. National Academy science letters, 40, 335-341. Available at: https://doi.org/10.1007/s40009-017-0590-1.

[32] Ibrahim, M. E., Ahmed, A. E., & Almujahed, H. (2019). A Comparative Study of Energy Saving Routing Protocols for Wireless Sensor Networks. International Journal on Information Technologies & Security, 11(2).

[33] Ibrahim, M. E., & Ahmed, A. E. (2022). Energy-aware intelligent hybrid routing protocol for wireless sensor networks. Concurrency and Computation: Practice and Experience, 34(3), e6601.

[34] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. Advances in engineering software, 69, 46-61. Available at: https://doi.org/10.1016/j.advengsoft.2013.12.007.

[35] Zheng, B., & Yang, J. (2021). Measurement method of distributed nodes in wireless sensor networks based on multiple attributes. Scientific Programming, 2021, 1-11. Available at: https://doi.org/10.1155/2021/9936337.

[36] Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on wireless communications, 1(4), 660-670. Available at: https://doi.org/10.1109/TWC.2002.804190.

[37] Abo-Zahhad, M., Ahmed, S. M., Sabor, N., & Sasaki, S. (2014). A new energy-efficient adaptive clustering protocol based on genetic algorithm for improving the lifetime and the stable period of wireless sensor networks. International Journal of Energy, Information and Communications, 5(3), 47-72.

[38] Kumar, R., & Kumar, D. (2016). Hybrid swarm intelligence energy efficient clustered routing algorithm for wireless sensor networks. Journal of sensors, 2016. Available at: https://doi.org/10.1155/2016/5836913

Research on Qubit Mapping Technique Based on Batch SWAP Optimization

Hui Li, Kai Lu, Zi’ao Han, Huiping Qin, Mingmei Ju, Shujuan Liu
School of Computer and Information Engineering, Harbin University of Commerce, Harbin, China

***Abstract*—The conventional approach for initial qubit mapping in the Noisy Intermediate-Scale Quantum (NISQ) era typically uses a static heuristic strategy, overlooking insufficient qubit neighborhood in subsequent operations, resulting in excess additional SWAP gates. To address this, we introduce a multifactor interaction cost function considering qubit distance, interaction time, and gate operation error rates, enhancing SWAP gate selection in the traditional strategy. Considering quantum hardware constraints, we propose Batch SWAP Optimization Strategy (BSOS). BSOS tackles qubit mapping challenges by leveraging optimal SWAP gate selection and a SWAP-based batch update technique, effectively minimizing SWAP gates throughout circuit execution. Experimental results show that BSOS significantly reduces additional gates by intelligently selecting SWAP gates and using batch updating, with a 38.1% average decrease in inserted SWAP gates, leading to a 12% reduction in hardware gate counting overhead.**

***Keywords*—*Quantum computing; quantum circuit compilation; initial qubit mapping; Batch SWAP Optimization Strategy* (*BSOS*)*; best SWAP choice; Batch Update Technology* (*BUT*)**

I. Introduction

Quantum computing, a revolutionary paradigm, has transformed finance [1], machine learning [2], optimization [3], and chemistry [4]. Traditional computers face challenges in complex problem solving and large-scale data processing due to resource limitations. Quantum computers offer a new solution with inherent parallelism. In quantum computing, Hamiltonian quantities describe problem evolution, translated for simulation using algorithms like Grover's search [5] and Shor's algorithm [6]. The product formula, approximating Hamiltonian exponentiation, is crucial. Quantum circuits excel in manipulating exponents, making them powerful for Hamiltonian simulation. Leveraging quantum circuit advantages efficiently advances quantum computing.

High-level representations of quantum circuits are not inherently tied to specific hardware and require translation into an instruction set compatible with the underlying quantum hardware for execution. In NISQ computers, the prevalent instruction set typically comprises a single-qubit rotation gate and one or more two-qubit gates. These quantum computers can only apply two-qubit gates to a limited set of qubit pairs due to constrained connections between qubits. Consequently, it becomes essential to perform circuit-to-hardware qubit mapping using initial qubit mapping techniques. Additionally, to scale up the circuit by increasing the number of gates and depth, it is necessary to reposition qubits to neighboring locations by introducing SWAP gates.

The existing literature on time scheduling strategies [7] offers solutions for compiling circuit depth, yet it may encounter challenges in managing constraints and optimization, particularly with intricate circuits. Another approach in [8], focusing on time scheduling and constraint planning, generates a hot-start solution but may face limitations regarding computational complexity and algorithmic efficiency, especially for large-scale circuits. A proposed greedy stochastic search approach [9] proves effective for similar problems but may struggle with complex circuits and global optimization. Genetic algorithms with chromosome coding strategies, introduced in [10] for optimization, might face challenges related to algorithmic parameter sensitivity and convergence speed. The study in [11] explores the trade-off between switched gates and circuit depth during compilation but offers limited consideration of hardware characteristics such as gate error rate. The research in [12] and [13] introduces strategies considering gate error rates, a significant advancement, yet practical applications may necessitate a more comprehensive consideration of hardware characteristics like cooling time and connectivity. In summary, while these literatures contribute valuable insights to the problem, further improvements and a more holistic approach may be required to address complex circuits and global optimization effectively.

In this paper, we explore Hamiltonian operator arrangements' flexibility, propose a multifactor interaction cost function and introduce BSOS for the initial qubit mapping process, aiming to efficiently compile quantum circuits for 2-local qubit Hamiltonian simulation problems. Given prevalent NISQ computer characteristics, where two-qubit gate error rates are typically 10 times higher than single-qubit gates, and qubit coherence time is shorter [14], BSOS adapts to diverse qubit topologies and gate sets. It seamlessly integrates with various quantum circuit mapping algorithms and is suitable for quantum approximation optimization algorithms like Quantum Approximate Optimization Algorithm (QAOA) [15]. Through evaluations, BSOS substantially reduces the required gate number compared to the state-of-the-art initial qubit mapping strategy. We further validate the effectiveness of BSOS through experiments on an IBM quantum device.

The main contributions of this paper can be summarized as follows:

- We focus on the initial qubit mapping phase, identifying the challenges and limitations that are the primary focus of this study.
- We introduce a multifactor interaction cost function, considering qubit distance, interaction time, and gate

error rate. This cost function facilitates a more comprehensive evaluation and optimization of quantum circuit performance.

- We propose an optimal SWAP gate selection algorithm and define SWAP gains. The SWAP gain is assessed based on the benefits obtained by adding SWAP gates. Optimal SWAP gates are selected for different instruction sets in quantum circuits.
- We design a scalable SWAP-based batch update technique, providing comparable results to previous mapping-based update-by-sequence approaches. This rapid update scheme ensures the scalability of qubit mapping, allowing the BSOS to adapt to various hardware architectures and accommodate larger quantum devices in NISQ era.

The subsequent sections of the paper are organized as follows: Section II presents relevant background information on quantum simulation and quantum circuits. Section III introduces the modification scheme for the cost function and details the BSOS strategy. The evaluation of these approaches is discussed in Section IV. Finally, the paper is summarized in Section V.

II. Preliminaries

A. Product Formula (Trotter's Formula)

In the realm of quantum computing, the product formula, also known as Trotter's formula, stands out as a fundamental technique for constructing efficient circuit structures. It leverages the decomposition of Hamiltonian quantities, representing the time evolution of a system in exponential form. The essence of the product formula lies in its ability to approximate time evolution by breaking down the system's Hamiltonian quantity, denoted as H, into distinct operators comprising sums of polynomial ergodic terms. Quantum circuits are then employed to efficiently implement these operators. The formula can be succinctly expressed as:

$$V(t) = \prod_{j=1}^{L} \exp(ith_j H_j) \tag{1}$$

where, $V(t)$ represents the state of the system at moment t, L is the number of terms, h_j is the coefficient of the jth term and H_j is the corresponding ergodic operator.

If all terms are exchangeable (i.e., $H_jH_k = H_kH_j$), then the product formula approximates the true time evolution U, (i.e., $V(t) = U$). However, in natural physical systems, non-exchangeable terms are typically present. In such cases, a first-order approximation can be used to approximate $V(t)$ as an rth power of $V(t/r)$, where r is a constant greater than 1. This process is known as the Trotterization step, and repeating this step r times forms a Trotter sequence. By decreasing the value of r, the cost of the simulation can be significantly reduced.

In this paper, we mainly consider the 2-local qubit Hamiltonian quantity, as shown in Eq. (2):

$$H = \sum_{(u,v)\in E} H_{uv} + \sum_{k\in V} H_k \tag{2}$$

where, H_{uv} represents a two-qubit Hamiltonian term and H_k is a single-qubit Hamiltonian. The interaction graph of this Hamiltonian quantity is represented by **G**(**V**, **E**), where **V** represents the set of qubits and **E** represents the set of edges.

B. Quantum Circuit

In quantum computing, data is stored in qubits, each of which has two fundamental states, denoted by $|0\rangle$ and $|1\rangle$. Unlike classical bits, qubits can be in a superposition of these two fundamental states, i.e., $\alpha\ |0\rangle + \beta\ |1\rangle$, where α and β are complex numbers and satisfy $|\ \alpha\ |^2 + |\ \beta\ |^2 = 1$.

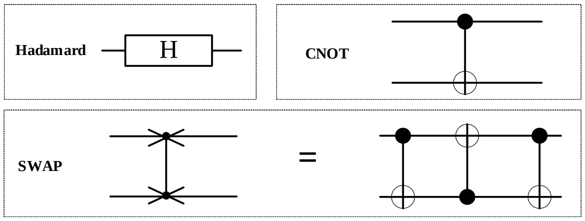


Fig. 1. Basic gates of IBM quantum computers.

Operations in quantum computing are realized through quantum gates, which apply specific operations like rotations, flips, etc., between qubits. Quantum gates are mathematically represented by unitary matrices. A Hadamard gate operates on a single qubit, while both CNOT gates and SWAP gates act on two qubits, as shown in Fig. 1. A CNOT gate flips the state of the target qubit based on the state of the controlling qubit, i.e., CNOT: $|c\rangle|t\rangle \rightarrow |c\rangle|c \oplus t\rangle$, where c, $t \in \{0, 1\}$ and $\oplus$ denotes a heteroskedastic operation. the SWAP gate then exchanges the states of the two target qubits: for all a, $b \in \{0, 1\}$, it will $|a\rangle|b\rangle \rightarrow |b\rangle|a\rangle$. The SWAP gate can be realized by a combination of 3 CNOT gates (see Fig. 1).

A quantum circuit is a framework for describing and manipulating quantum information, comprising a sequence of quantum gates akin to classical logic gates. Quantum gates enact transformations on quantum states. Quantum circuits can be conceptualized as systems constructed from a combination of fundamental quantum gates and quantum measurements. In the realm of intricate quantum operations, like simulating Hamiltonian quantities, quantum gates within quantum circuits can be deliberately designed and fine-tuned.

The quantum mapping task involves a graph **G = (V, E)** that represents the structure of the target quantum device and a circuit **C** representing an ideal quantum algorithm. The gates in circuit **C** are decomposed into elementary gates supported by the target quantum device. The objective of quantum mapping is to transform circuit **C** into a functionally equivalent circuit. In this transformed circuit, each two-qubit gate acts on a pair of neighboring nodes in the graph **G** of the target quantum device. Essentially, the goal is to adapt and map the circuit **C** according to the physical architecture of the target quantum device, ensuring its compatibility with the specific quantum hardware. Fig. 2(a) shows an example of a circuit where **Q** = $\{q_0, ..., q_4\}$, **C** = $\{g_0, ..., g_6\}$, where g_0 = CNOT (q_0, q_1), g_1 = CNOT (q_2, q_4), and so on. In the figure, each CNOT gate is labelled with the qubit they act on. In the text, q *is* used to represent logical qubits and P is used to represent physical qubits.

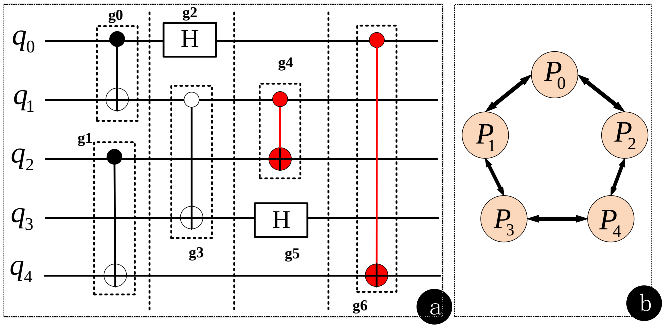


Fig. 2. (a) A logical circuit with depth 4, (b) Architecture diagram of quantum device.

Circuit **C** is typically represented as a sequence of gates (g_0, g_1, ..., g_{m-1}), but this cannot imply that in all cases the (i + s)th gate has to be executed after the *ith* gate (where $i \geq 0$, $s \geq 1$, and $0 < i + s < m$). In reality, if the two gates don't involve common qubits, they can be executed in parallel. For instance, considering the circuit in Fig. 2(a), it can be expressed as:

$$C = \left(\langle q_0, q_1 \rangle, \langle q_2, q_4 \rangle, \langle q_0 \rangle, \langle q_1, q_3 \rangle, \langle q_1, q_2 \rangle, \langle q_3 \rangle, \langle q_0, q_4 \rangle \right).$$

III. Proposed Methodologies

A. Problem in Initial Mapping

Qubit mapping aims to determine the optimal arrangement of qubits, minimizing the number of qubit shift operations needed for all two-qubit gates. Like the methodologies outlined in [16]-[19], the qubit mapping problem is cast as a Quadratic Assignment Problem (QAP).

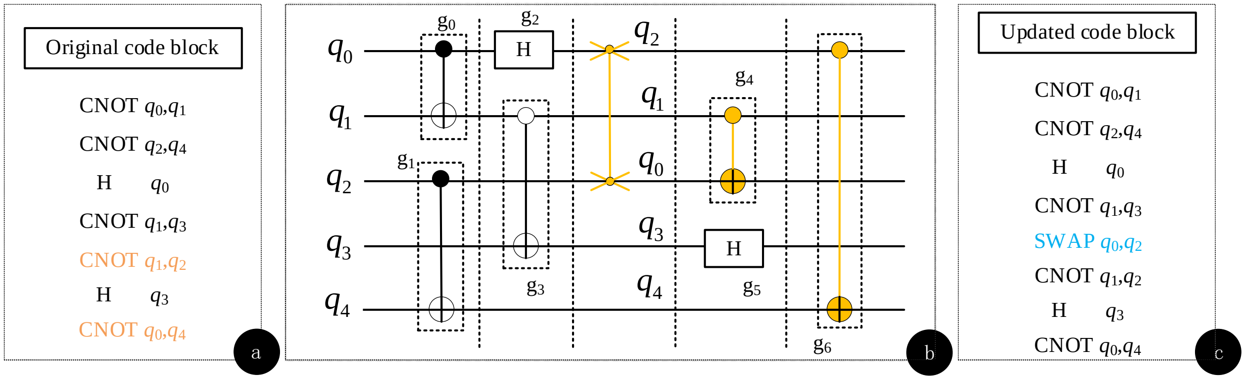
Similar to the previous approach, a SWAP operation is employed to alter the state between two qubits, facilitating the adjustment of qubit mapping. Introducing 1 SWAP gate increases the circuit's depth by 3. Multiple swap gates enable the relocation of a logical qubit to any physical qubit position. Fig. 3 illustrates that after inserting a SWAP operation between q_0 and q_2 following the third CNOT gate, the modified quantum circuit becomes executable. The first 3 CNOT gates can be executed under the initial qubit mapping, and after inserting the SWAP, the mapping is updated to $\{q_0 \rightarrow p_2, q_1 \rightarrow p_1, q_2 \rightarrow p_0, q_3 \rightarrow p_3, q_4 \rightarrow p_4\}$, and the remaining two CNOT gates can now be executed under this updated mapping.

Comparing the initial and optimized circuits in Fig. 2(a) and Fig. 3, it is evident that the number of gates increases from 7 to 10, and the circuit depth increases from 4 to 7. The introduction of additional SWAP gates notably enhances the circuit's execution time. Hence, the primary objective of the mapping process is to minimize the number of additional SWAP gates inserted, aiming to reduce the overall error rate and total execution time of the final hardware-adapted circuit.

Definition 1 Qubit mapping [6]: given a coupling map of input quantum circuits and quantum devices, find the initial qubit mapping and the intermediate qubit mapping transformations (by insertion swapping) to satisfy all two qubit constraints and try to minimize the number of additional gates and the circuit depth in the final hardware-compatible circuit.

In the initial qubit mapping phase of quantum circuit compilation, several challenges need to be addressed. This paper considers the following limitations:

- Selection of physical qubits: It is crucial to choose the appropriate physical qubits to map the logical qubits. Taking into account the hardware's topology, physical qubits are selected to minimize communication overhead.
- Connectivity limitation: Some hardware platforms only permit direct communication between specific qubits, while other communications need to be achieved through SWAP gates.
- SWAP gate cost: The execution cost of SWAP gates is typically high, so the cost of SWAP gates is taken into account when choosing the initial qubit mapping scheme, with a preference for less costly SWAP operations.
- Optimal performance trade-off: The goal of the initial qubit mapping is to achieve efficient quantum gate operations while minimizing the communication overhead. Different performance metrics such as communication overhead, SWAP gate cost, etc., need to be weighed when choosing the physical qubits and order.


Original code block

CNOT q_0,q_1
CNOT q_2,q_4
H q_0
CNOT q_1,q_3
CNOT q_1,q_2
H q_3
CNOT q_0,q_4

Updated code block

CNOT q_0,q_1
CNOT q_2,q_4
H q_0
CNOT q_1,q_3
SWAP q_0,q_2
CNOT q_1,q_2
H q_3
CNOT q_0,q_4

Fig. 3. (a) Original code block, (b) Updated hardware-compliant quantum circuit, (c) Updated code block.

Improving the quality of the initial qubit mapping requires addressing the following questions:

- How to determine the appropriate mapping targets? This involves identifying, for a selected gate, which qubits are the targets. In other words, it determines which qubits need to be communicated or exchanged during the execution of the gate. Once the best mapping target has been selected, the appropriate mapping operation can be performed, such as the execution of a SWAP gate to exchange the positions of qubits. This ensures that the target operation can be executed correctly in hardware.
- How to efficiently choose where to insert SWAP gates, and how many to insert, in order to improve the mapping and minimize the total number of SWAP gates.

B. Design the Cost Function

In the exploration of potential SWAP gates, the role of the cost function is to prioritize the SWAP gate that closely aligns with the target mapping in terms of cost. Assessing potential SWAP gates entails the selection of the most favorable SWAP sequence, constituting an ongoing decision-making process integral to the overall mapping strategy. The arrangement requires reevaluation when updating the qubit mapping following each SWAP operation, constituting a dynamic process. To achieve optimal mapping outcomes, the mapping must be recalibrated and adjusted subsequent to each SWAP insertion through the utilization of a cost function.

Among numerous mapping methods, the cost function serves as a pivotal tool for assessing the effectiveness of a mapping strategy. However, prevailing cost functions typically focus on a singular factor, predominantly relying on distance as the fundamental metric—whether physical or logical distance between qubits. These functions often lack the capability to consider multiple factors. Physical distance impacts the operational speed and error rate, as qubits situated farther apart may necessitate more steps and involve intermediate bits with a higher probability of errors during operations.

To address the aforementioned issue, this paper introduces a multifactor interaction cost function as shown in Eq. (3). This function incorporates three primary factors: distance, interaction time, and error rate of gate operation. Minimizing the cost function enables the identification of the most efficient and accurate strategies and configurations for accomplishing a specific quantum computing task. This cost function is

$$Y_{\mathrm{cos}t}(E, f_{ij}, d_{\phi(i)\phi(j)}) = \min_{\phi \in S_n} \sum_{i=1}^{n} \sum_{j=1}^{n} E f_{ij} (d_{\phi(i)\phi(j)})^2 \quad (3)$$

where, S_n denotes all possible permutations, f_{ij} represents the interaction time between logic qubits i and j in the circuit, and the interaction time can be estimated based on hardware specifications or empirical experiments. e denotes the probability of an error occurring for the execution of a double quantum gate, where $0 \leq E \leq 1$, and $d_{\varphi(i)\varphi(j)}$ denotes the physical distance between the qubits $\varphi(i)$ and $\varphi(j)$ on the hardware, which can be calculated by using the Floyd-Warshall algorithm to perform the calculation.

The physical distance between qubits plays a pivotal role in influencing the speed and efficacy of their interactions. Increased distances directly contribute to heightened computational complexity and error rates, with the square factor of distance exerting a substantial impact on overall computational cost, thereby influencing the efficiency and accuracy of quantum computation. Interaction time serves as a determinant of task execution speed, with prolonged interaction times resulting in extended task execution durations. Introducing interaction time as a factor in the cost function underscores its significance in determining the overall computational cost and efficiency. Quantum gate operations inherently incur error rates, where elevated error rates undermine computational accuracy and reliability. The incorporation of gate operation error rates into the cost function underscores the critical influence of accuracy and reliability on the cost and efficiency of quantum computing. By amalgamating these three factors, the initial qubit mapping of a quantum circuit can be generated with a more comprehensive consideration of hardware limitations and practical implementation constraints.

C. Best SWAP Choice

In the realm of quantum computing, any multi-quantum gate can be decomposed into a combination of single quantum gates and CNOT gates. The execution of these gates necessitates the inclusion of SWAP gates to alter the connectivity pattern between qubits. To minimize the surplus of added SWAP gates, strategic selection of SWAP gates that can yield more nearest-neighbor (NN) gates is crucial. This practice contributes to the implementation of intricate quantum operations and algorithms, enhancing the coherence of the circuit. The significance of SWAP gates lies in two primary aspects: Firstly, SWAP gates facilitate the interchange of states between two qubits, thereby reshaping the relationships between qubits and optimizing the structure and efficiency of the quantum circuit. Secondly, in practical quantum operations, achieving the target superposition quantum state often involves constructing a combination of corresponding quantum gates. A higher count of NN gates enhances the likelihood of attaining this objective. Formally, SWAP gate gains are defined as follows:

Definition 2 SWAP gate gains: let P be the set of non-nearest-neighbor (non-NN) gates, Q be the set of executable SWAP gates, $\mathrm{V}_{\mathrm{SWAP}}$ = **{P, Q}** be the list of SWAP gains, N_1 be the set of executable $\mathbf{V_{SWAP}}$, i.e., the number of NN gates after adding this SWAP gate. N_0 is the number of NN gates when the $\mathbf{V_{SWAP}}$ set is not executed. The optimal benefit $\mathbf{V_{SWAP}}$ set is determined by selecting max $\{N_1 - N_0\}$.

$$\mathrm{V}_{\mathrm{SWAP}} = N_1 - N_0 \quad (4)$$

We introduce a heuristic-based strategy designed to optimize the selection of SWAP gates, with the goal of minimizing the compilation overhead in quantum circuits. The strategy involves simulating the execution of a SWAP operation for each potential candidate, resulting in a new mapping. Subsequently, the sizes of the elements in VSWAP are compared, determined by the number of NN gates generated under the new mapping with different SWAP gates inserted. Among all feasible SWAP operations, the one with the largest element size is chosen. This approach enables a more intelligent selection of SWAP gates, thereby reducing the overall count of SWAP gates and enhancing the overall performance of the quantum circuit.

Algorithm 1 Best SWAP Choice

Input.
nn_gate_count: The number of nearest-neighbor gates in the circuit after adding a swap gate between two qubits
moves: A set of insertable SWAP gates with the same cost
Output.
best_move: Optimal insertable SWAP gate

```
begin
    max_nn_gate_increase ← 0
    best_move ← None
    for move in moves do
        nn_gate_increase ← nn_gate_count[move]
            if nn_gate_increase > max_nn_gate_increase
                max_nn_gate_increase ← nn_gate_increase
                best_move ← move
        end if
    end for
end
```

Details of the pseudo-code can be found in Algorithm 1 in the text, while a specific application example is provided in Fig. 4.

Compile an 8-qubit 2-local Hamiltonian into the lattice architecture depicted in Fig. 4. In the presence of a set of insertable SWAP gates with identical costs, the approach outlined in this paper is to identify the one with the highest gain among these gates. Fig. 4(a) presents a qubit map of a circuit along with a scenario where specific quantum operations algorithms cannot be implemented without the insertion of SWAP gates. The upper figure displays the inserted SWAP gates and the CNOT gates implemented after insertion, while the lower figure illustrates the qubit graph, where nodes represent qubits, and edges signify their connectivity. To prevent confusion, the SWAP gates in the figure are applied to the corresponding hardware qubits. For enhanced readability, they are plotted on the circuit qubits.

In the circuit, CNOT (0, 1) and CNOT (1, 5) are direct mappings, but SWAP gates are needed to perform the remaining CNOT gates. SWAP gates with the same cost are calculated according to Eq. (3), such as SWAP (2, 5) and SWAP (5, 6) in figure. Subsequently, the filtered SWAP gates are inserted into the circuits separately to see the number of NN gates generated under the new mapping. As in Fig. 4(b) SWAP (q_2, q_5) is inserted and V_{SWAP} = 2 under the current mapping (see Eq. 4). Whereas in Fig. 4(c), V_{SWAP} = 1 after inserting SWAP (q_5, q_6) By comparison, a set of SWAP gates with a larger number of NN gates is selected and inserted into the circuit to update the mapping. And so on until all the quantum gates are mapped and the final mapping result is shown in Fig. 4(d). This strategy ensures that the selected SWAP gates maximize the number of NN gates and optimize the mapping of the quantum circuit.

D. Batch Update Technology

The traditional approach to quantum mapping involves real-time updates of mappings to adjust the relationship after each SWAP operation. However, as quantum circuits increase in size, this method becomes inefficient. To address this issue, we propose a Batched Update Technique (BUT).

The design of the BUT is based on reducing the cost associated with frequent mapping updates in conventional quantum mapping methods. The strategy aims to enhance the efficiency of quantum circuit mapping by decreasing the frequency of updates, incorporating the overall circuit structure, and balancing mapping performance with the associated update costs, among other theoretical considerations. The fundamental concept is to holistically consider multiple mapping update operations to reduce interference with circuit execution operations and achieve a trade-off between mapping quality and update costs. Specifically, each batch update takes into account all relevant SWAP operations and executes mapping updates for these operations simultaneously. This approach reduces the number of mapping updates stemming from a single SWAP operation, thereby diminishing latency and energy consumption in quantum computing.

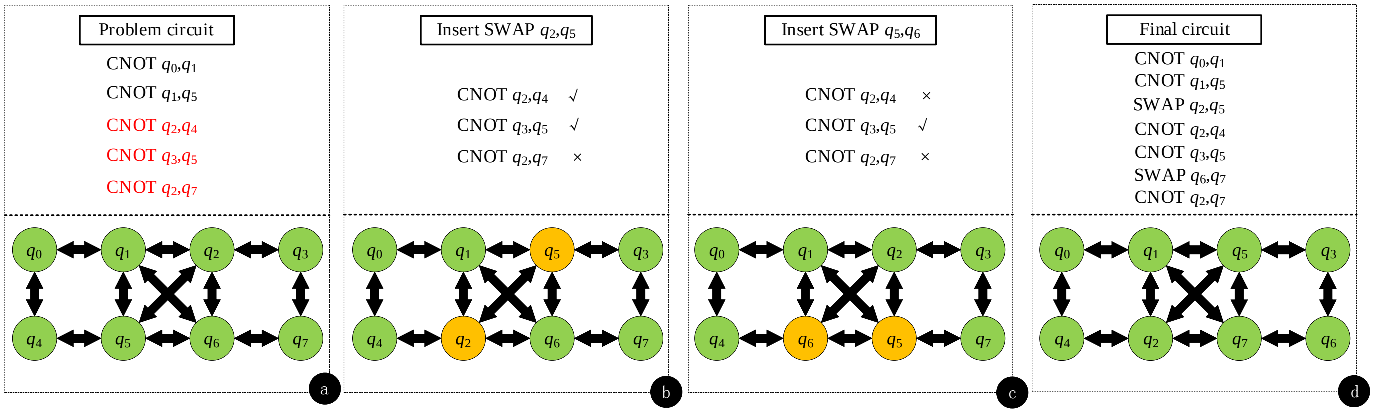
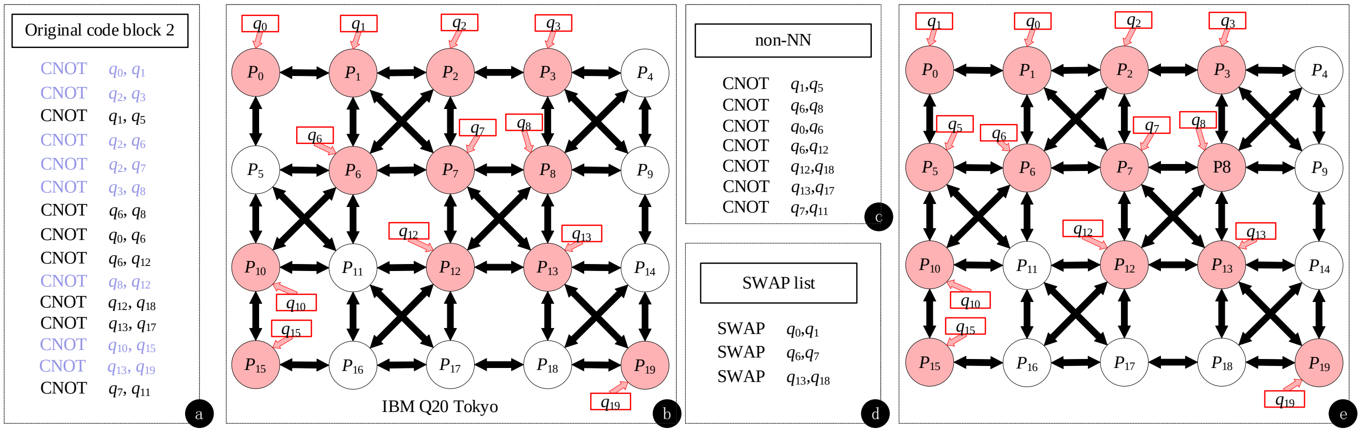


Fig. 4. Examples of compiling a 8-qubit 2-local Hamiltonian to a grid architecture. (a) A problem circuit, (b-c) Insert SWAP. (b) The NN gate that can be realized after inserting SWAP (q_2, q_5), (c) The NN gate that can be realized after inserting SWAP (q_5, q_6), (d) Final circuit.



```
Original code block 2

CNOT    q0, q1
CNOT    q2, q3
CNOT    q1, q5
CNOT    q2, q6
CNOT    q2, q7
CNOT    q3, q8
CNOT    q6, q8
CNOT    q0, q6
CNOT    q6, q12
CNOT    q8, q12
CNOT    q12, q18
CNOT    q13, q17
CNOT    q10, q15
CNOT    q13, q19
CNOT    q7, q11
```

```
non-NN

CNOT    q1,q5
CNOT    q6,q8
CNOT    q0,q6
CNOT    q6,q12
CNOT    q12,q18
CNOT    q13,q17
CNOT    q7,q11
```

```
SWAP list

SWAP    q0,q1
SWAP    q6,q7
SWAP    q13,q18
```

Fig. 5. (a) Original code block 2, (b) Coupling graph of IBM q20, (c) non-NN quantum gate, (d) List of SWAP gates to be executed, (e) Batch update mapping results.

In extensive quantum circuits, the interaction constraints between qubits necessitate multiple SWAP operations to facilitate the exchange between non-adjacent qubits. Each SWAP operation triggers a mapping update, contributing to inefficiency. Consequently, consolidating multiple operations and updating the mapping collectively after completing all operations emerges as an appealing solution. The BUT scrutinizes the entire quantum circuit to identify which two-qubit gates can be optimized through a shared SWAP operation. When a SWAP gate is chosen for execution, it gets added to a "list of SWAP operations to be performed," with a defined condition dictating when to cease additions, such as reaching a predetermined list length. During batch execution of SWAP operations, the algorithm iterates through the list, executes all listed SWAP operations, and updates the qubit mapping collectively upon completion.

Fig. 5 shows an example, assuming the circuit shown in Fig. 5(a) is run on the 20-qubits device Tokyo (Fig. 5(b)). First, two-qubit gates that are NNs on the initial qubit layout, such as those labelled purple in Fig. 5(a), are filtered from the original circuit list2. Map these gates directly to the corresponding hardware (e.g., the qubits labelled pink in Fig. 5(b)), i.e.

$q_0 \rightarrow p_0$, $q_1 \rightarrow$ p$_1$, $q_2 \rightarrow p_2$, $q_3 \rightarrow p_3$, $q_6 \rightarrow p_6$, $q_7 \rightarrow p_7$, $q_8 \rightarrow p_8$, $q_{10} \rightarrow p_{10}$, $q_{12} \rightarrow p_{12}$, $q_{13} \rightarrow p_{13}$, $q_{15} \rightarrow p_{15}$, $q_{19} \rightarrow p_{19}$.

For non-NN two-qubit gates, the costs between qubits are compared by means of a computed cost function. The qubit pair with the smallest cost is chosen as the target of the mapping, assuming that the SWAP gate with the smallest cost is evaluated as $\{q_0, q_1\}$. The previous method is to add that SWAP gate to the execution list, update the qubit mapping, and remove the NN gates from the unmapped set of gates. The steps are repeated until all double qubit gates are mapped. Instead of executing the SWAP gates directly, the strategy in this paper puts the SWAP (q_0, q_1) into the SWAP list (e.g., Fig. 5(d)), and then searches for the next SWAP gate, which is also added to the list. When the length of the list reaches a predefined value, it traverses the "list of SWAP operations to be performed" and then performs all SWAP operations in the list. After all the above SWAP operations are completed, the qubit mapping is updated uniformly.

In this process, handling multiple operations simultaneously may introduce some complexity, as it is necessary to ensure that there are no conflicts between batch operations and to update the qubit mapping correctly. Whenever a new SWAP operation is added to the "pending SWAP operation list," a conflict check is performed for this operation with other operations already in the list. Specifically, it is ensured that the new SWAP operation does not impact or be impacted by the SWAP operations already in the list. The qubit mapping is then updated in bulk based on the selected SWAP list, i.e., when each new SWAP operation is added to the "pending SWAP operation list.", i.e.

$q_0 \rightarrow p_1$, $q_1 \rightarrow p_0$, $q_2 \rightarrow p_2$, $q_3 \rightarrow p_3$, $q_5 \rightarrow p_5$, $q_6 \rightarrow p_7$, $q_7 \rightarrow p_6$, $q_8 \rightarrow p_8$, $q_{10} \rightarrow p_{10}$, $q_{11} \rightarrow p_{11}$, $q_{12} \rightarrow p_{12}$, $q_{13} \rightarrow p_{18}$, $q_{15} \rightarrow p_{15}$, $q_{17} \rightarrow p_{17}$, $q_{18} \rightarrow p_{13}$, and $q_{19} \rightarrow p_{14}$

Remove the NN from the unmapped gates and empty the "list of pending SWAP operations" in preparation for the next batch of SWAP operations. Repeat until all double-qubit gates are mapped.

IV. Result and Discussion

Similar to earlier studies, this paper uses the following metrics to evaluate the performance of different compilers: the aggregate count of inserted SWAP gates (lower values are preferable) and the overall count of executed two-qubit gates on the hardware (lower values are preferable). These metrics enable the evaluation of algorithm performance, particularly in handling intricate circuits. The benchmarking procedure aligns with IBM's Qiskit quantum program, utilizing the Qiskit compiler for decomposition and optimization of the CX/CNOT gate set. The benchmarking methodology outlined in [13], focusing on the QAOA model, was adopted, wherein the time evolution of Hamiltonian quantities is conducted by multiplying Eq. (5) by:

$$V(t) = (\Pi_{j=1}^{L} \exp(ih_j H_j t / r))^r \tag{5}$$

where, r is the number of iterations of Trotter. The coefficients of H_j were randomly selected in the range (0, π). The evaluation ranges from 4 to 22 qubits, running their mapping process five times and selecting the best result.

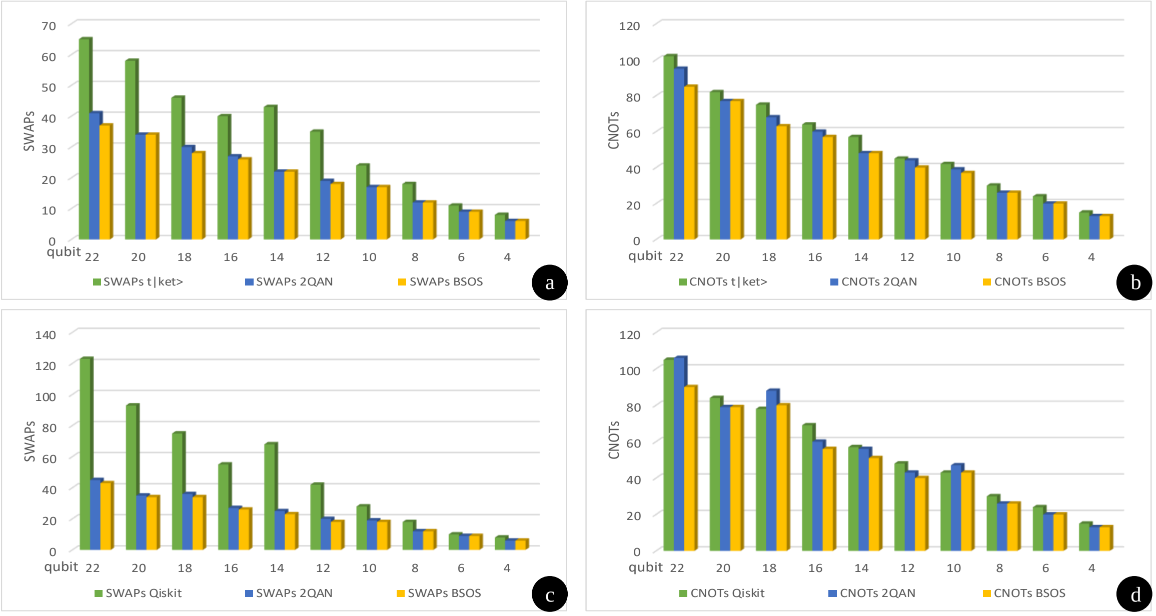


Fig. 6. (a) Comparison of SWAP gate compilation cost between BSOS technology and t|ket⟩ and 2QAN, (b) Comparison of CNOT gate compilation cost between BSOS technology, t|ket⟩ and 2QAN, (c) Comparison between BSOS technology and Qiskit and Comparison of SWAP gate compilation cost of 2QAN, (d) Comparison of BSOS technology, t|ket⟩ and 2QAN, (e) Comparison of SWAP gate compilation cost of 2QAN (c) Comparison between BSOS technology and Qiskit and Comparison of SWAP gate compilation cost of 2QAN, (d) Comparison of BSOS technology with CNOT gate compilation cost of Qiskit and 2QAN.

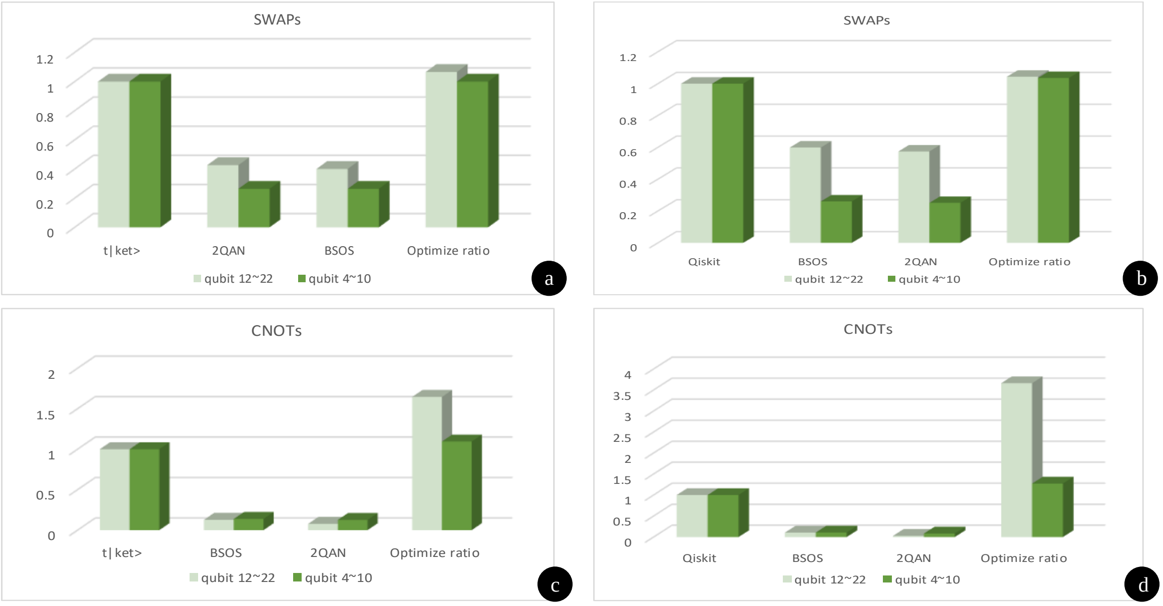


Fig. 7. The BSOS strategy was evaluated for circuits with 4 to 10 qubits and 12 to 22 qubits. (a) SWAP gate optimization rate of BSOS relative to t|ket⟩ and 2QAN, (b) SWAP gate optimization rate of BSOS relative to Qiskit and 2QAN, (c) CNOT gate optimization rate of BSOS relative to t|ket⟩ and (d) CNOT gate optimization rate of BSOS relative to t|ket⟩ and 2QAN. optimization rate of BSOS relative to Qiskit and 2QANs

BSOS was implemented in Python 3.9 and all compilations were performed on a laptop with an Intel Core i5 processor (2.30GHz and 8GB RAM).

Compare the BSOS compilation strategy with the compilation overheads of t|ket⟩ and Qiskit. Fig. 6 and Fig. 7 show the compilation results on the IBM. Compared to t|ket⟩ and Qiskit, BSOS has the least compilation overhead in terms of the number of SWAP gates inserted, the number of hardware dual-quantum gates, and the circuit depth. Specifically, the t|ket⟩ compiler [20] (version 0.11.0) and the Qiskit compiler [21] (version 0.26.2, optimization level 3), equipped with the recommended "FullPass", are considered.

The IBM quantum compiler is limited to CNOT or CZ gate sets, so the models were evaluated using the compilation results on the IBM quantum computer. Of all the benchmarks and quantum computers, the run using the BSOS strategy in the QAP mapping turned out to be the best, as can also be seen in Fig. 6, where the optimization is more pronounced with a higher number of qubits. For 22 qubits, BSOS inserts 43% less SWAP counts than t|ket⟩ and 65% less than Qiskit (Fig. 6(a)(b)). This reduction in SWAP count will lead to a reduction in the number of hardware double-qubit gates, and BSOS reduces the double-qubit gate overhead by 17% and 14% (Fig. 6(c)(d)).

In the t|ket⟩ compiler, for the case of 4 to 10 qubits, the average reduction in the number of inserted SWAP gates is 26.4%, while the average reduction in the number of inserted CNOT gates is 13.8%. This optimization effect is 10% higher than using only 2QAN. After optimization for 12 to 22 qubits, the average reduction in the number of inserted SWAP gates across all evaluated benchmarks is 40%, and the average reduction in the number of inserted CNOT gates is 12.8%. This optimization effect is 70% higher than using only 2QAN (see Fig. 7(a) (b)).

In the Qiskit compiler, for 4 to 10 qubits, the average reduction in the number of inserted SWAP gates is 26%, while the average reduction in the number of inserted CNOT gates is 10.8%. This optimization effect is 30% higher than using only 2QAN. For 12 to 22 qubits, after optimization, the average reduction in the number of inserted SWAP gates is 60%, and the average reduction in the number of inserted CNOT gates is 10.6% (see Fig. 7(c) (d)).

In this study, optimizing the placement of qubits is taken as the main concern. To solve this problem efficiently, the Tabu search algorithm was chosen. This algorithm is very fast in solving small scale problems, e.g., for the QAOA model with 4 qubits, it takes only about 0.221 seconds in the t|ket⟩ compiler and about 0.004 seconds in the Qiskit compiler, as shown in Table I. However, the processing speed drops significantly when faced with problems of larger size. For example, the QAOA model with 20 qubits takes about 24.218 seconds in the t|ket⟩ compiler and about 0.0176 seconds in the Qiskit compiler.

By applying this cost function, quantum computing researchers and engineers can more accurately quantify the impact of different designs and strategies on the system performance, providing a scientific basis for decision-making and advancing the development of quantum computing technology. In order to verify the accuracy and effectiveness of the proposed cost function, a comparison experiment is conducted in the t|ket⟩ compiler for the distance-only cost function and the cost function proposed in this paper, and the results are shown in Fig. 8. From the figure, it can be seen that the compilation result of the cost function in this paper is better. For 12~22 bits, SWAP is reduced by 28.52% and CONT is reduced by 35.81% on average. While for 4~10 bits, SWAP is reduced by 7.32% and CONT is reduced by 13.74% on average.

TABLE I. COMPARING AVERAGE RUNNING TIMES OF 2QAN AND BUT IN COMPILERS T|KET⟩ AND QISKIT

qubit	Running time		
		BUT	2QAN
22	t|ket⟩	22.218	22.553
	Qiskit	0.0176	0.0184
20	t|ket⟩	10.274	10.954
	Qiskit	0.0178	0.0178
6	t|ket⟩	0.318	0.361
	Qiskit	0.005	0.006
4	t|ket⟩	0.221	0.244
	Qiskit	0.004	0.004

The superior performance of BSOS in the 22-qubit scenario primarily manifests in its intelligent SWAP gate selection strategy, the comprehensive consideration using a multifactor interaction cost function, and the introduction of batch update techniques. These advantages enable BSOS to more effectively optimize the placement of quantum bits in large-scale quantum circuits. Specifically, the intelligent SWAP gate selection and comprehensive consideration of multiple factors enhance the overall mapping performance, while the batch update technique reduces mapping costs. These optimization effects are particularly pronounced in the case of 22 qubits.

The BSOS algorithm not only demonstrates outstanding performance in the 22-qubit scenario but also holds broad potential applications and future research directions. The application areas of BSOS include the compilation and execution of large-scale quantum circuits, especially well-suited for highly optimized quantum tasks. Future research directions may encompass optimizing nested quantum algorithms, adapting the BSOS algorithm to dynamic scenarios, deeper integration of quantum hardware characteristics, and incorporating BSOS into comprehensive quantum compilation automation tools, providing support for the further development of quantum computing technology.

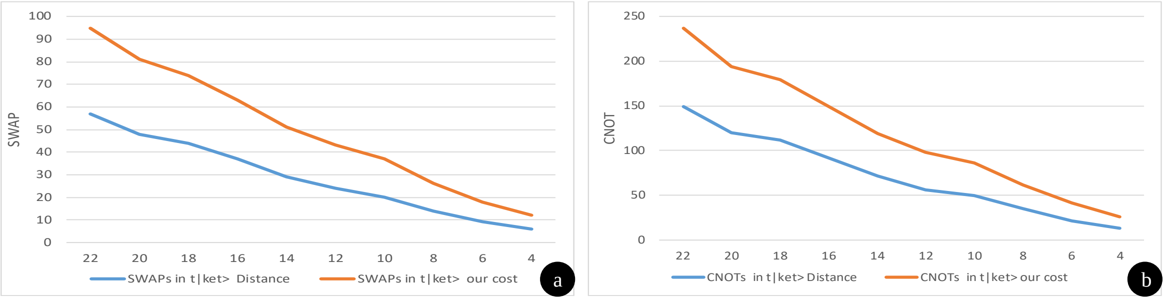


Fig. 8. (a) The number of SWAP gates under different cost functions, (b) The number of CNOT gates under different cost functions.

V. Conclusion

In the NISQ era, there is still a significant gap between quantum software and imperfect NISQ hardware. This research introduces a Bulk SWAP Optimization Strategy (BSOS) specifically designed for addressing the 2-local qubit Hamiltonian simulation problem. Focusing on the adaptable operators within Hamiltonian quantities, the primary optimization targets the initial qubit mapping of qubits. A comprehensive evaluation reveals that the BSOS strategy significantly mitigates compilation overhead on the IBM quantum computer, demonstrating superior performance compared to the other two general-purpose compilers.

The optimal SWAP gate selection algorithm optimizes circuit locality by selecting SWAP gates that generate a larger number of newly added NN gates, while the SWAP updating strategy reduces the frequency of mapping by batch updating and optimizing the timing, which improves the overall efficiency of quantum circuit mapping. On the other hand, the introduction of qubit interaction time and the error rate of gate operation in the cost function helps to improve the efficiency and reliability of quantum computation, which makes up for the lack of comprehensiveness and accuracy of previous methods. This makes the proposed method more applicable to NISQ computers with different characteristics and optimization goals, and provides a useful improvement direction for the efficient execution of mesoscale quantum computation. Looking ahead, more optimization work is planned and other possible research directions are explored. By applying error mitigation techniques, it is expected that the error rate can be further reduced, thus further improving the performance and reliability of quantum computation.

Acknowledgment

This work was partially supported by the Natural Science Foundation in Heilongjiang Province of China, under Grant LH2022F035; in part by the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province, under Grant UNPYSCT-2020212; and in part by the Science Foundation of Harbin Commerce University, under Grant XL0095.

References

[1] Stamatopoulos, N., Egger, D. J., Sun, Y., Zoufal, C., Iten, R., Shen, N., & Woerner, S. (2020). Option pricing using quantum computers. *Quantum*, *4*, 291.

[2] Zoufal, C., Lucchi, A., & Woerner, S. (2019). Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Information*, *5*(1), 103.

[3] Harwood, S., Gambella, C., Trenev, D., Simonetto, A., Bernal, D., & Greenberg, D. (2021). Formulating and solving routing problems on quantum computers. *IEEE Transactions on Quantum Engineering*, *2*, 1-17.

[4] Jang, S. J. (2023). Quantum Mechanics for Chemistry. *Quantum*.

[5] Grover, L. K. (1996, July). A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (pp. 212-219).

[6] Shor, P. W. (1999). Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, *41*(2), 303-332.

[7] Cotta, C., & Fernández, A. J. (2007). Memetic algorithms in planning, scheduling, and timetabling. In *Evolutionary Scheduling* (pp. 1-30). Berlin, Heidelberg: Springer Berlin Heidelberg.

[8] Booth, K., Do, M., Beck, J., Rieffel, E., Venturelli, D., & Frank, J. (2018, June). Comparing and integrating constraint programming and temporal planning for quantum circuit compilation. In *Proceedings of the International Conference on Automated Planning and Scheduling* (Vol. 28, pp. 366-374).

[9] Oddi, A., & Rasconi, R. (2018). Greedy randomized search for scalable compilation of quantum circuits. In *Integration of Constraint Programming, Artificial Intelligence, and Operations Research: 15th International Conference, CPAIOR 2018, Delft, The Netherlands, June 26–29, 2018, Proceedings 15* (pp. 446-461). Springer International Publishing.

[10] Rasconi, R., & Oddi, A. (2019, July). An innovative genetic algorithm for the quantum circuit compilation problem. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 7707-7714).

[11] Li, G., Ding, Y., & Xie, Y. (2019, April). Tackling the qubit mapping problem for NISQ-era quantum devices. In *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems* (pp. 1001-1014).

[12] Tannu, S. S., & Qureshi, M. K. (2018). A case for variability-aware policies for nisq-era quantum computers. *arXiv preprint arXiv:1805.10224*.

[13] Murali, P., Baker, J. M., Javadi-Abhari, A., Chong, F. T., & Martonosi, M. (2019, April). Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers. In *Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems* (pp. 1015-1029).

[14] Li, S., Nguyen, K. D., Clare, Z., & Feng, Y. (2023, October). Single-Qubit Gates Matter for Optimising Quantum Circuit Depth in Qubit Mapping. In *2023 IEEE/ACM International Conference on Computer Aided Design (ICCAD)* (pp. 1-9). IEEE.

[15] Tate, R., Farhadi, M., Herold, C., Mohler, G., & Gupta, S. (2023). Bridging classical and quantum with SDP initialized warm-starts for QAOA. *ACM Transactions on Quantum Computing*, *4*(2), 1-39.

[16] Dousti, M. J., Shafaei, A., & Pedram, M. (2014, May). Squash: a scalable quantum mapper considering ancilla sharing. In *Proceedings of the 24th edition of the great lakes symposium on VLSI* (pp. 117-122).

[17] Bahreini, T., & Mohammadzadeh, N. (2015). An MINLP model for scheduling and placement of quantum circuits with a heuristic solution approach. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, *12*(3), 1-20.

[18] Lao, L., van Wee, B., Ashraf, I., van Someren, J., Khammassi, N., Bertels, K., & Almudever, C. G. (2018). Mapping of lattice surgery-based quantum circuits on surface code architectures. *Quantum Science and Technology*, *4*(1), 015005.

[19] Lao, L., & Browne, D. E. (2022, June). 2qan: A quantum compiler for 2-local qubit hamiltonian simulation algorithms. In *Proceedings of the 49th Annual International Symposium on Computer Architecture* (pp. 351-365).

[20] Sivarajah, S., Dilkes, S., Cowtan, A., Simmons, W., Edgington, A., & Duncan, R. (2020). t|ket⟩: a retargetable compiler for NISQ devices. *Quantum Science and Technology*, *6*(1), 014003.

[21] Carrazza, S., Efthymiou, S., Lazzarin, M., & Pasquale, A. (2023, February). An open-source modular framework for quantum computing. In *Journal of Physics: Conference Series* (Vol. 2438, No. 1, p. 012148). IOP Publishing.

Identifying Factors in Congenital Heart Disease Transition using Fuzzy DEMATEL

Raghavendra M Devadas[1], Vani Hiremani[2*], Ranjeet Vasant Bidwe[3], Bhushan Zope[4], Veena Jadhav[5], Rohini Jadhav[6]
Department of Computer Science and Engineering, Gitam School of Technology, GITAM Bengaluru, India[1]
Symbiosis Institute of Technology, Symbiosis International (Deemed University) (SIU), Lavale, Pune, India[2, 3, 4]
Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune[5, 6]

Abstract—**The transition from pediatric to adult cardiology care is a pivotal moment in the healthcare journey of individuals with congenital heart conditions or childhood-onset heart diseases. This multifaceted process requires meticulous consideration of clinical, psychosocial, and logistical factors. This research aims to explore the critical criteria for transitioning pediatric patients to adult cardiology, delving into the challenges and opportunities inherent in this healthcare shift. The identified factors for successful transition, including age and developmental stage, medical complexity, cardiac function, psychosocial factors, insurance, and financial considerations, play integral roles in the transition process. Leveraging analytical methodologies, particularly the Fuzzy Decision-Making Trial and Evaluation Laboratory (DEMATEL), this study involves three experts who assess criteria linguistically, converted to Triangular Fuzzy Numbers, and averaged. Defuzzification, using the CFCS method, yields crisp values. Results reveal that Medical Complexity (U+V = 3.96, U-V = 0.233), Insurance (U+V = 3.931, U-V = 0.22), Psychosocial Factors (U+V = 3.839, U-V = 0.387), and Age and Developmental Stage (U+V = 3.802, U-V = 0.106) follow Cardiac Function (U+V = 4.312, U-V = 0.946) in ranking. Age and Developmental Stage, Medical Complexity, Psychosocial Factors, and Insurance are considered causal variables, with Cardiac Function as an effect. These numerical insights enhance our understanding of transition criteria interdependencies, informing tailored healthcare strategies.**

Keywords*—*DEMATEL; Fuzzy DEMATEL; factors; pediatric patients; heart disease

I. Introduction

Congenital Heart Disease (CHD) is a complex and diverse group of cardiovascular conditions that affect a substantial number of children worldwide. The remarkable advancements in medical care, surgical interventions, and early diagnosis have significantly improved the survival rates of pediatric patients born with CHD. As a result, more and more of these individuals are now moving from receiving care in cardiology to adult cardiology as they enter adolescence and early adulthood. This shift, in healthcare can be sometimes difficult phase for them. It's essential to transition patients with congenital heart disease (CHD) into adult cardiology care to ensure that they continue to receive effective treatment, monitoring and support. However, this process is complex. Its outcome has an impact, on the long-term health and quality of life of these patients.

The transition covers aspects, such, as medical considerations, psychological and social factors, communication and involving both patients and their families. This study builds on the work of researchers [1, 2] a review by [3] and the guidelines provided in the literature to further understand how pediatric patients with heart disease move from pediatric care to adult care. By using the Fuzzy DEMATEL approach our goal is to identify the factors that influence this transition. This information can then be used by healthcare providers, policymakers and stakeholders to improve the quality of care and long-term outcomes, for this group of patients. As we strive for a patient centered transition model, the importance of utilizing tools has become more recognized. This research aims to address the urgency of identifying and prioritizing the key determinants that influence the transition of CHD pediatric patients into adolescents. This research makes a significant contribution to the field of pediatric cardiology by applying the Fuzzy DEMATEL approach to unravel the intricate dynamics of transitioning pediatric patients with CHD to adolescence. By utilizing this advanced analytical tool, the study aims to provide a nuanced understanding of the critical factors influencing the transition process. This study's contribution lies in its potential to identify, quantify, and prioritize the key elements that impact the successful transition of pediatric CHD patients to adult care. The outcomes are expected to go beyond traditional analyses, offering insights into the fuzzy relationships among various transition criteria. Such insights can inform tailored strategies for healthcare providers, policymakers, and stakeholders to optimize the transition process. These findings can be used to develop more effective transition procedures and strategies, thereby improving the quality and outcomes of care for this vulnerable population. The outcomes of this study hold the potential to transform the landscape of pediatric-to-adolescent transition care, offering hope and improved prospects for those living with congenital heart disease. This paper is organized as follows: Literature review is presented in Section II, Section III discusses scaffold of methodologies, in Section IV discussion about the results is made and finally Section V concludes the paper.

II. Literature Review

The Decision-making Trial and Evaluation Laboratory (DEMATEL) method is a widely used multicriteria decision-making method [4]. It is used to evaluate the relationships between factors in various fields [5]. DEMATEL helps to extract the complex structure of a problem by identifying cause-and-effect relationships among different elements [6]. It is used to model the understandable structure of a complex system and measure the complexity of a problem [7]. DEMATEL can be applied to both small and simple systems as well as complex systems [8]. The aim of [9] is to provide a solution to the issues mentioned above. Several barriers impede blockchain adoption, including system-related, external, intra-organizational, and inter-organizational ones. The research in [10] looks at the obstacles that stand in the way when it comes to assessing blockchain life cycles in China. There has been no research to scrutinize how these barriers cooperate to improve decision-making in life cycle assessments. In a study by [11] identify critical bottlenecks in the development of the HRS in China using a customized Fuzzy DEMATEL method. It is suggested by the authors of [12], that a Fuzzy trapezoidal approach should be employed to prioritize software requirements. Various Fuzzy logic-based approaches for prioritizing software requirements are discussed in [13]. The study in [14] suggests a new version of the fuzzy DEMATEL that uses PFS for language variables. The research in [15] focuses on establishing a logical framework for strategy maps; inserting subjectivity into the overall strategy formulation process; identifying the most important connections between strategy objectives to make the route map clear, useful, and easy to understand; and combining qualitative and quantitative methods to make a strong and complete solution.

III. Theoretical Background

A. Fuzzy Logic

Fuzzy logic is a mathematical framework that contracts with vagueness besides imprecision in decision-making and control systems. It is an extension of binary logic that is based on true and false. It was first proposed in the early 1960s by the Iranian mathematician Lotfi Zadeh. Fuzzy logic allows for degrees of truth, which makes it particularly useful in situations where information is vague, ambiguous, or incomplete.

Definition 1: A triangular fuzzy number, also known as a TFN is defined by a membership function that assigns degrees of membership (degrees of truth) to values, within a given range. The membership function of a triangular number takes the shape of a triangle and determined by three parameters; the lower bound (a) the upper bound (b) and the peak (c). We can express a triangular number using the equation:

$$\begin{array}{ll} \mu\_A(x) = \{\ 0, & \text{if } x < u, \\ \quad (x - u) / (w - u & \text{if } u \leq x < w, \\ \quad (v - x) / (v - w), & \text{if } w \leq x < v, \\ \quad 0, & \text{if } x \geq v \\ \quad \} & \end{array} \tag{1}$$

where, $\mu\_A(x)$ is the degree of membership (or membership value) of value x to the triangular fuzzy number A. Where u is a lower bound, v an upper one, and w is the peak. This function defines the degree of membership for any provided value x within [u, v].

Definition 2: When working with two Triangular Fuzzy Numbers (TFNs) we perform operations to combine or manipulate these sets. Let us take two TFNs, A and B and their membership functions defined as follows:

For TFN A:

$$\begin{array}{ll} \mu\_A(x) = \{0, & \text{if } x < u1, \\ \quad (x - u1) / (c1 - u1), & \text{if } u1 \leq x < w1, \\ \quad (v1 - x) / (v1 - w1), & \text{if } w1 \leq x \leq v1, \\ \quad 0, & \text{if } x > v1 \\ \quad \} & \end{array}$$

For TFN B:

$$\begin{array}{ll} \mu\_B(x) = \{0, & \text{if } x < u2, \\ \quad (x - u2) / (w2 - u2), & \text{if } u2 \leqslant x < w2, \\ \quad (v2 - x) / (v2 - w2), & \text{if } w2 \leqslant x \leqslant v2, \\ \quad 0, & \text{if } x > b2 \\ \quad \} & \end{array}$$

1) *Basic operations:*

a) Addition of TFNs (U + V): To add two TFNs together you we add their membership values pointwise. The resulting TFN, denoted as C (U + V) can be represented by the following membership function:

$$\mu\_W(x) = \max(\mu\_U(x) + \mu\_V(x)) \tag{2}$$

where, $\mu\_W(x)$ represents the degree of membership of x to the resulting TFN W.

b) Subtraction of TFNs (U - V): The subtraction of two TFNs is performed by subtracting the membership values of the second TFN (V) from the membership values of the first TFN (U). The resulting TFN, W (U - V), has a membership function as follows:

$$\mu\_W(x) = \max(\mu\_U(x) - \mu\_V(x)) \tag{3}$$

*c) Multiplication of TFNs (U * V):* Multiplication of two TFNs is done by taking the product of their membership values at each point. Let W (U * V) be the resulting TFN function as follows:

$$\mu\_W(x) = \mu\_U(x) * \mu\_V(x) \tag{4}$$

d) Division of TFNs (U / V): The division of two TFNs is performed by dividing their membership values pointwise. The resulting TFN, W (U / V), has a membership function (for $\mu\_V(x) > 0$) as follows:

$$\mu\_W(x) = \mu\_U(x) / \mu\_V(x) \tag{5}$$

B. DEMATEL Method

The DEMATEL methodology provides an approach, for analyzing the cause-and-effect relationship of factors in complex decision-making problems [16]. It aims to understand how these factors are interconnected and their impact, on each other. This methodology is commonly employed in management, engineering and social sciences.

1) *DEMATEL Steps:*

a) Constructing a Causal Diagram (Impact Matrix): First, a causal diagram or an impact matrix is constructed for DEMATEL. The criterion measured is represented by the rows and columns of this matrix. The components of the matrix, labeled a_ij, depict the effect of directionality (or impact) of element i upon element j. The impact values can be determined based on expert opinions, surveys, or data analysis.

b) Normalization of the Impact Matrix: To ensure that the impact values are within a common scale, the impact matrix remains normalized. This is done by dividing each row of the matrix by the sum of its absolute values. The normalized matrix is denoted as N.

$$Z\_ij = |a\_ij| / \Sigma|a\_ij| \quad (\forall j) \tag{6}$$

where, Z_ij stands the normalized impact value of factor i on factor j.

c) Calculation of Total Influences: DEMATEL calculates the total influence of each factor by summing the normalized values in the corresponding row of the normalized impact matrix as shown in the below equation.

$$TI\_i = \Sigma Z\_ij \quad (\forall j) \tag{7}$$

where, TI_i is the total influence of factor i.

d) Division into cause-and-effect groups: Grounded on the total influence values, factors are categorized into two clusters: cause and effect. Since these were found to affect each other directly, and hence interchangeably we can say all had ‘‘cause’’ and ‘‘effect” attributes.

e) Drawing the Causal Diagram: A causal diagram is created to visually represent the cause-and-effect relationships among factors. This diagram shows how factors influence each other and helps in understanding the structure of the problem.

f) Interpreting the Results: The DEMATEL results are interpreted to determine the key elements, that plays an important role in the issue at consideration. Additionally, the method provides insights into the direction besides the strength of the causal associations.

C. Fuzzy DEMATEL (FDEMATEL) Method

FDM is an extension of the traditional DEMATEL technique that incorporates the concepts of fuzzy logic to handle vagueness besides imprecision in decision-making and problem-solving. DEMATEL is a technique employed in the analysis and visualization of the causal relationships between various elements in intricate systems. FDM, therefore, adds a layer of fuzziness to these relationships to better reflect real-world situations where relationships are not always clear-cut. The assessment of key determinants of congenital cardiac disease in young adults during the transitional period from adolescence to adulthood requires a collaborative decision-making approach. By interacting among various experts, a satisfactory decision is reached in a collective decision. Humans commonly criticize based on their experiences and insights in such a group decision-making process. As such, these decisions are made in an uncertain situation, and their expressions are more likely to be ambiguous than crisp. Consequently, the DEMATEL method cannot directly identify critical factors under these circumstances, requiring an extended DEMATEL method based on fuzzy set theory.

Step 1: The Fuzzy Direct Relation Matrix (FDRM) generation

To figure out how the n criteria relate to each other, we first make a matrix called n × n. Each element in every row in the matrix affects the element in each column in the matrix, and this can be represented by a fuzzy number. In the event of a matrix composed of the opinions of multiple experts, all experts must complete the matrix; the arithmetic mean of the opinions of all experts is used to construct the matrix z.

$$z = \begin{bmatrix} 0 & \cdots & \tilde{z}_{n1} \\ \vdots & \ddots & \vdots \\ \tilde{z}_{1n} & \cdots & 0 \end{bmatrix} \tag{8}$$

Step 2: The fuzzy direct-relation matrix normalization

The formula shown in Eq. (9) is used to compute the normalized fuzzy direct-relation matrix.

$$\tilde{a}_{ij} = \frac{\tilde{z}_{ij}}{r} = \left(\frac{m_{ij}}{r}, \frac{n_{ij}}{r}, \frac{o_{ij}}{r}\right) \tag{9}$$

Where,

$$r = \max_{i,j}\left\{\max_i \sum_{j=1}^{n} o_{ij}\, , \max_j \sum_{i=1}^{n} o_{ij}\right\} \qquad i, j \in \{1,2,3,\ldots,n\}$$

Step 3: The fuzzy total-relation matrix (TRM) is computed by the following formula:

$$\widetilde{TR} = \lim_{k\to+\infty}(\tilde{a}^1 \oplus \tilde{a}^2 \oplus \ldots \oplus \tilde{a}^k) \tag{10}$$

In the fuzzy total-relation matrix, the function of each element is expressed as $\tilde{t}_{\mathrm{ij}} = (\mathrm{m}^{"}_{\mathrm{ij}}, \mathrm{n}^{"}_{\mathrm{ij}}, \mathrm{o}^{"}_{\mathrm{ij}})$ and is generated as,

$$[m^{"}_{ij}] = a_m \times (I - a_m)^{-1}$$

$$[n^{"}_{ij}] = a_n \times (I - a_n)^{-1}$$

$$[o^{"}_{ij}] = a_o \times (I - a_o)^{-1}$$

In the case of a normalized matrix, the initial step is to identify the inverse of the standardized matrix. Subsequently, the inverse of matrix I is extracted and multiplied by the normalized matrix.

Step 4: Defuzzification for Crisp Values Generation

A crisp value of the TRM has been obtained using the CFCS. The CFCS method is as follows:

$$m^n_{ij} = \frac{(m^t_{ij} - min\, m^t_{ij})}{\Delta^{max}_{min}} \tag{11}$$

$$n^n_{ij} = \frac{(n^t_{ij} - min\, m^t_{ij})}{\Delta^{max}_{min}} \tag{12}$$

$$o^n_{ij} = \frac{(o^t_{ij} - min\, m^t_{ij})}{\Delta^{max}_{min}} \tag{13}$$

So that,

$$\Delta^{max}_{min} = \max o^t_{ij} - \min m^t_{ij} \tag{14}$$

Computing the normalized upper and lower bounds values:

$$m^s_{ij} = \frac{n^n_{ij}}{(1 + n^n_{ij} - m^n_{ij})} \tag{15}$$

$$o_{ij}^{s} = \frac{o_{ij}^{n}}{(1+ o_{ij}^{n} - m_{ij}^{n})} \tag{16}$$

The crisp value is the result of the CFCS algorithm.

The normalized crisp values:

$$x_{ij} = \frac{[l_{ij}^{s}(1-l_{ij}^{s})+u_{ij}^{s}\times u_{ij}^{s}]}{[1-l_{ij}^{s}+u_{ij}^{s}]} \tag{17}$$

Step 5: Threshold value

The threshold values are crucial in determining the internal relations matrix. As a result, NRMs are plotted by disregarding the partial relationships are disregarded. In the NRM, only those relationships with greater values than a threshold value in matrix TR are included. To calculate a threshold value for a given relationship, it suffices to calculate its average values from the matrix TR. Once the threshold intensity has been determined, all the values of matrix TR that are less than the threshold are set to 0, i.e., the causal relationship mentioned above is disregarded. For example, the matrix TR value is disregarded if the threshold value is 0.3970.

Step 6: Final result besides forming a causal relation diagram.

The subsequent step is to calculate the total of the rows and columns of TR (Step 4). The total of the rows (U) and columns (V) can be derived as follows:

$$U = \sum_{j=1}^{n} TR_{ij} \tag{18}$$

$$V = \sum_{i=1}^{n} TR_{ij} \tag{19}$$

Subsequently, the U and V determine the values of U+V and U-V.

Empirical Study: The fuzzy DEMATEL approach was used to extract all the dominant factors while transitioning from congenital heart disease to adulthood.

Step 1: The study identified five important dominant factors as tabulated in Table I. The study considers three experts providing a direct impact on each of the dominant factors in terms of linguistic assessment, the linguistic terminology with their respective TFNs is depicted in Table II.

Step 2: DRMs provided by experts

The linguistic terminology mentioned by all three experts are converted into corresponding TFNs are tabulated in Table III to Table V and the arithmetic mean of all the three Tables is shown in Table VI.

Step 3: Normalize the Fuzzy DRM as per Eq. (9) and the normalized DRM is tabulated as shown in Table VII. Also, the fuzzy TRM is calculated as the Eq. (10) and is shown in Table VIII.

Step 4: Defuzzify into crisp values and threshold values: Total normalized crisp values are calculated as per Eq. (17) and are shown in Table IX. As mentioned in Step 5 of Section 3.3 the threshold value for this study is set to 0.3970. Table X describes the model of important relations.

TABLE I. DOMINANT FACTORS (DF)

Factor	Explanation
DF1	Age and Developmental Stage: Determine the appropriate age for transition, considering both chronological age and developmental maturity. Some patients may be ready for transition earlier or later based on their circumstances
DF2	Medical Complexity: Evaluate the complexity of the patient's heart condition. Patients with complex congenital heart defects or ongoing medical issues may require specialized adult cardiology care
DF3	Cardiac Function: Assess the current cardiac function and stability of the patient. This includes evaluating factors like ejection fraction, valve function, and the need for ongoing interventions or surgeries.
DF4	Psychosocial Factors: Consider the patient's psychosocial well-being and readiness for transition. Assess their understanding of their condition, self-management skills, and emotional preparedness for adult care.
DF5	Insurance: Address insurance coverage and financial considerations. Verify that the patient's insurance plan will cover adult cardiology care and that there are no disruptions in coverage during the transition

TABLE II. LINGUISTIC TERMINOLOGY WITH THEIR TFNS

Linguistic terms	M	N	O
No influence	1	1	1
Very low influence	2	3	4
Low influence	4	5	6
High influence	6	7	8
Very high influence	8	9	9

TABLE III. FIRST EXPERT DRM

	DF1	DF2	DF3	DF4	DF5
DF1	0	2	5	3	3
DF2	2	0	4	3	4
DF3	2	2	0	3	4
DF4	5	2	3	0	4
DF5	1	5	5	3	0

TABLE IV. SECOND EXPERT DRM

	DF1	DF2	DF3	DF4	DF5
DF1	0	2	4	2	2
DF2	3	0	5	3	4
DF3	1	3	0	4	2
DF4	4	5	4	0	3
DF5	4	3	5	2	0

TABLE V. THIRD EXPERT DRM

	DF1	DF2	DF3	DF4	DF5
DF1	0	4	4	5	2
DF2	3	0	5	2	3
DF3	5	3	0	2	1
DF4	3	1	3	0	4
DF5	3	4	5	1	0

Step 5: Output and CR diagram

As per Eq. (18) and Eq. (19) the concluding output is computed and shown in Table XI.

The model of significant relationships is illustrated in Fig. 1, which can be visualized as a graph, with the values of (U+V) and (U-V) positioned on the horizontal and vertical axes respectively. The coordinate system determines the positions and relationships of each factor to a given point in the coordinate system.

IV. RESULTS AND DISCUSSION

By Fig. 1 and Table XI, each DF was assessed based on the following elements:

- The U + V horizontal vector indicates the importance of each factor within the system. More specifically, it indicates the effect of factor I within the system and the effect of other system factors within the system on the factor. Cardiac Function is the ranking factor in terms of importance, followed by Medical Complexity, insurance, psychosocial factors, age and development stage, and insurance. Age and development stage are considered causal variables in this study, while Cardiac Function is considered an effect.
- The vertical vector U-V represents the extent to which a factor influences a system. In cases a positive U-V indicates the cause while a negative U-V suggests the effect. In this study cardiac function emerges as the factor determining importance followed by factors, like medical complexity, insurance, psychosocial aspects, age and development stage, in descending order of significance.

Table VII presents the Normalized Fuzzy DRM, a decision-making matrix that has been normalized to facilitate a comprehensive analysis of the relationships between the identified factors (DF1, DF2, DF3, DF4, and DF5). The entries in the matrix are represented as fuzzy numbers, denoted by (lower bound, mean, upper bound), capturing the uncertainty and imprecision in the decision-making process. The fuzzy numbers in each cell represent the normalized influence of the corresponding row factor on the column factor. The lower and upper bounds capture the range of possible influence, while the mean value provides a central tendency. Table VIII extends the analysis by presenting the Fuzzy Total Relation, which aggregates the normalized influences from Table VII to provide a holistic view of the overall relationships between the factors.

Table IX presents the crisp TRM, a triangular fuzzy relationship matrix, for the identified factors in the study. The matrix is symmetric, with each cell indicating the degree of relationship strength between two factors. the values within the matrix range between 0 and 1, representing the intensity of the relationship. Higher values indicate stronger relationships, while lower values suggest weaker relationships. The diagonal elements of the matrix are typically 1, indicating the self-relationship of each factor. This matrix provides a quantitative representation of the fuzzy relationships among the factors under consideration.

TABLE VI. ARITHMETIC MEAN OF EXPERT DRMS

	DF1	**DF2**	**DF3**	**DF4**	**DF5**
DF1	(0.00,0.00,0.00)	(3.33,4.33,5.33)	(6.66,7.66,8.33)	(4.66,5.66,6.33)	(2.66,3.66,4.66)
DF2	(3.33,4.33,5.33)	(0.00,0.00,0.00)	(7.33,8.33,8.67)	(3.33,4.33,5.33)	(5.33,6.33,7.33)
DF3	(3.67,4.33,4.67)	(3.33,4.33,5.33)	(0.00,0.00,0.00)	(4.00,5.00,6.00)	(3.00,3.66,4.33)
DF4	(6.00,7.00,7.67)	(3.67,4.33,4.67)	(4.67,5.67,6.67)	(0.00,0.00,0.00)	(5.33,6.33,7.33)
DF5	(3.67,4.33,5.00)	(6.00,7.00,7.67)	(8.00,9.00,9.00)	(2.33,3.00,3.67)	(0.00,0.00,0.00)

TABLE VII. NORMALIZED FUZZY DRM

	DF1	**DF2**	**DF3**	**DF4**	**DF5**
DF1	(0.00,0.00,0.00)	(0.10,0.13,0.16)	(0.20,0.23,0.25)	(0.14,0.17,0.19)	(0.08,0.11,0.14)
DF2	(0.10,0.13,0.16)	(0.00,0.00,0.00)	(0.22,0.25,0.26)	(0.10,0.13,0.16)	(0.16,0.19,0.22)
DF3	(0.11,0.13,0.14)	(0.10,0.13,0.16)	(0.00,0.00,0.00)	(0.12,0.15,0.18)	(0.09,0.11,0.13)
DF4	(0.18,0.21,0.23)	(0.11,0.13,0.14)	(0.14,0.17,0.20)	(0.00,0.00,0.00)	(0.16,0.19,0.22)
DF5	(0.11,0.13,0.15)	(0.18,0.21,0.23)	(0.24,0.27,0.27)	(0.07,0.09,0.11)	(0.00,0.00,0.00)

TABLE VIII. FUZZY TOTAL RELATION

	DF1	**DF2**	**DF3**	**DF4**	**DF5**
DF1	(0.13,0.24,0.43)	(0.21,0.36,0.58)	(0.36,0.54,0.79)	(0.24,0.38,0.58)	(0.20,0.34,0.57)
DF2	(0.23,0.38,0.60)	(0.14,0.26,0.47)	(0.40,0.59,0.84)	(0.21,0.36,0.58)	(0.27,0.42,0.61)
DF3	(0.21,0.32,0.49)	(0.19,0.32,0.51)	(0.16,0.30,0.50)	(0.20,0.32,0.51)	(0.19,0.31,0.50)
DF4	(0.30,0.44,0.65)	(0.24,0.38,0.59)	(0.34,0.54,0.80)	(0.13,0.25,0.44)	(0.28,0.42,0.65)
DF5	(0.24,0.37,0.57)	(0.30,0.44,0.64)	(0.42,0.60,0.82)	(0.19,0.33,0.53)	(0.14,0.26,0.45)

TABLE IX. THE CRISP TRM

	DF1	DF2	DF3	DF4	DF5
DF1	0.27	0.378	0.554	0.388	0.364
DF2	0.394	0.293	0.596	0.376	0.437
DF3	0.342	0.343	0.326	0.342	0.33
DF4	0.452	0.4	0.548	0.274	0.439
DF5	0.389	0.449	0.605	0.347	0.285

TABLE X. THE CRISP TRM INCLUSIVE OF THE THRESHOLD VALUE

	DF1	DF2	DF3	DF4	DF5
DF1	0	0	0.554	0	0
DF2	0	0	0.596	0	0.437
DF3	0	0	0	0	0
DF4	0.452	0.4	0.548	0	0.439
DF5	0	0.449	0.605	0	0

Table X extends the Crisp TRM by incorporating a threshold value to highlight significant relationships. In this modified matrix, values below the threshold are set to 0, indicating negligible or weak relationships, while values equal to or above the threshold are retained. This thresholding process simplifies the matrix by emphasizing only the most impactful relationships, making it easier to interpret and focus on the key interactions.

TABLE XI. FINAL RESULT

	V	U	U+V	U-V
Age and Developmental Stage	1.848	1.954	3.802	0.106
Medical Complexity	1.863	2.096	3.96	0.233
Cardiac Function	2.629	1.683	4.312	0.946
Psychosocial Factors	1.726	2.113	3.839	0.387
Insurance	1.856	2.075	3.931	0.22

Table XI presents the final results of the study, providing a comprehensive assessment of the identified factors in the context of the transition from pediatric patients with congenital heart disease to adolescence. The Table includes four columns: V, U, U+V, U-V.

Age and Developmental Stage: This factor is assigned values for each of the four metrics: V (1.848), U (1.954), U+V (3.802), U-V (0.106).

Medical Complexity: Similarly, this factor is assessed with values for V (1.863), U (2.096), U+V (3.96), U-V (0.233).

Cardiac Function: The values for this factor are V (2.629), U (1.683), U+V (4.312), U-V (0.946).

Psychosocial Factors: This factor is evaluated with values for V (1.726), U (2.113), U+V (3.839), U-V (0.387).

Insurance: The final results for this factor are V (1.856), U (2.075), U+V (3.931), U-V (0.22).

The results in Table XI offer a quantified understanding of the factors' individual contributions and their combined effects. These metrics provide insights into the central tendency, range, and upper bounds of the factors, allowing for a nuanced interpretation of their significance in the transition process.

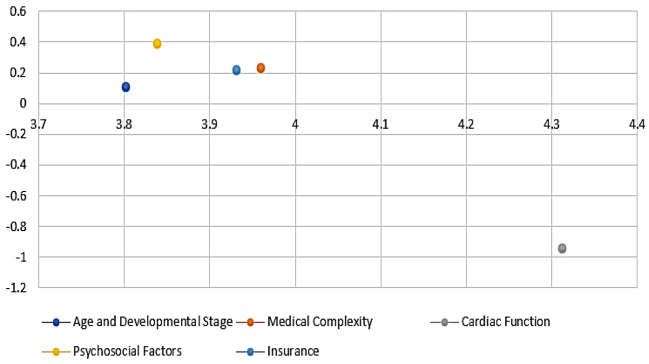


Fig. 1. Cause-Effect Graph.

The above Fig. 1 shows the relationship between five variables: Age and Developmental Stage, Medical Complexity, Psychosocial Factors, Insurance, and Cardiac Function. The x-axis reflects variable prominence, indicating influence levels, while the y-axis denotes their causative or resultant nature. Variables with high x-values and low y-values, like Cardiac Function and Insurance, are primary outcomes strongly influenced by others. Those with high x-values and near-zero y-values, such as Age and Developmental Stage, Medical Complexity, and Psychosocial Factors, act as both causes and effects. Variables with low x-values and near-zero y-values are independent, lacking strong influences. Notably, no variable has both high x and y-values, suggesting a lack of variables solely causing the problem.

V. CONCLUSION

In conclusion, the application of the Fuzzy DEMATEL approach has offered a quantified understanding of critical factors in the transition from pediatric patients with congenital heart disease to adolescence. The numerical values from Table XI underscore the distinct contributions of each factor, revealing noteworthy insights. Age and Developmental Stage, with a combined upper bound and mean value of 3.802, exhibits a significant overall influence with a relatively narrow range (U-V = 0.106). Medical Complexity demonstrates substantial impact (U+V = 3.96) with a moderate range (U-V = 0.233). Cardiac Function emerges as a key determinant (U+V = 4.312) with a broader impact range (U-V = 0.946). Psychosocial Factors and Insurance, with U+V values of 3.839 and 3.931, respectively, demonstrate moderate to substantial influences. These numerical assessments provide a tangible basis for prioritizing interventions and tailoring transition strategies, laying the groundwork for informed clinical and policy decisions in pediatric-to-adolescent transition care for congenital heart disease patients.

REFERENCES

[1] Mackie, Andrew S et al. “Transition Intervention for Adolescents With Congenital Heart Disease.” Journal of the American College of Cardiology vol. 71,16, pp.1768-1777, 2018.

[2] Dimopoulos, Konstantinos, et al. "Transition to adult care in adolescents with congenital heart disease." Progress in Pediatric Cardiology, vol. 51, pp. 62-66, 2018.

[3] Ozsahin, Dilber Uzun, et al. "Fuzzy logic in medicine." Biomedical Signal Processing and Artificial Intelligence in Healthcare, 153-182, 2020.

[4] Khatun, Marzana, et al. "An application of DEMATEL and fuzzy DEMATEL to evaluate the interaction of safety management system and cybersecurity management system in automated vehicles." Engineering Applications of Artificial Intelligence, vol. 124, pp. 106566, 2023.

[5] Šmidovnik, Tjaša, and Petra Grošelj. "Solution for Convergence Problem in DEMATEL Method: DEMATEL of Finite Sum of Influences." *Symmetry*, vol. 15, no. 7, pp. 1357, 2023.

[6] Tp, Krishnakantha et al. “Using this DEMATEL Corporate social responsibility CSR.” REST Journal on Banking, Accounting and Business, vol. 4, no. 4, n. pag, 2018.

[7] Du, Yuan-Wei, and Xin-Lu Shen. "Group hierarchical DEMATEL method for reaching consensus." *Computers & Industrial Engineering*, vol. 175, pp. 108842, 2023.

[8] Hassan, Muhammad I., et al. "The Dematel Method for Assessing Contributing Factors in University Selection." International Journal of Academic Research in Progressive Education and Development, vol. 11, no. 4, 2022.

[9] Li, Hai, et al. "A novel hybrid MCDM model for machine tool selection using fuzzy DEMATEL, entropy weighting and later defuzzification VIKOR." Applied Soft Computing, vol. 91, pp. 106207, 2020.

[10] Chen, Zhihua, et al. "Sustainable supplier selection for smart supply chain considering internal and external uncertainty: An integrated rough-fuzzy approach." Applied Soft Computing, vol. 87, pp. 106004, 2020.

[11] Yazdi, Mohammad, et al. "A novel extension of DEMATEL approach for probabilistic safety analysis in process systems." Safety Science, vol. 121, pp. 119-136, 2020.

[12] Devadas, Raghavendra, and Nagaraj G. Cholli. "PUGH Decision Trapezoidal Fuzzy and Gradient Reinforce Deep Learning for Large Scale Requirement Prioritization." Indian Journal of Science and Technology, vol. 15, no. 12, pp. 542-553, 2022.

[13] Devadas, Raghavendra and G. N. Srinivasan. “Review Of Different Fuzzy Logic Approaches for Prioritizing Software Requirements.” International Journal of Scientific & Technology Research. Vol. 8, no 09, pp. 296-298, 2019.

[14] Abdullah, Lazim, and Pinxin Goh. "Decision making method based on Pythagorean fuzzy sets and its application to solid waste management." Complex & Intelligent Systems, vol. 5, no. 2, pp. 185-198, 2019.

[15] Acuña-Carvajal, Felipe, et al. "An integrated method to plan, structure and validate a business strategy using fuzzy DEMATEL and the balanced scorecard." Expert Systems with Applications, vol. 122, pp. 351-368, 2019.

[16] Fontela, E. and Gabus, A. “The DEMATEL Observer”. Battelle Geneva Research Center, Geneva, 1976.

Reliable and Efficient Model for Water Quality Prediction and Forecasting

Azween Abdullah[1], Himakshi Chaturvedi[2], Siddhesh Fuladi[3],
Nandhika Jhansi Ravuri[4], Deepa Natesan[5], M.K Nallakaruppan[6]
Faculty of Applied Science and Technology, Perdana University, Malaysia[1]
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India-632014[2, 3, 4]
Department of Networking and Communication, SRM Institute of Science and Technology, Kattankulathur[5]
School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, India-632014[6]

***Abstract*—Water quality is a crucial aspect of environmental and public health. Hence, its assessment is of paramount importance. This research paper aims to leverage machine learning models to classify water quality based on a comprehensive dataset. The dataset contains various water quality indicators, and the primary objective is to predict whether the water is safe or not to consume or use. This research evaluates the performance of diverse machine learning algorithms, such as Decision Trees, Random Forest, Logistic Regression, Support Vector Machines, and more for comparative analysis. Performance metrics such as accuracy, precision, recall, and F1-score are used to assess the models' effectiveness in classifying water quality. The Random Forest algorithm gave the best performance with an accuracy of 95.08%, an F1-Score of 94.69%, a Precision of 90.48%, a Recall of 93.10%, and an AUC score of 0.91. A comparative plot for the ROC AUC curve is also plotted between the various machine learning models used. Feature importance, which can help identify which water quality parameters have the greatest impact on predicting water quality outcomes, is also found in the research work.**

***Keywords*—Random forest; logistic regression; feature importance; decision trees; support vector machines**

I. Introduction

Access to clean and safe drinking water is a fundamental human right. Waterborne diseases resulting from contaminated water sources have severe consequences on public health. Water quality plays a pivotal role in ensuring the well-being of both ecosystems and human populations. However, despite international efforts to ensure safe water sources for all, the global challenge of providing clean and potable water persists. Hence, monitoring water quality is essential to prevent waterborne diseases and environmental degradation. The World Health Organization (WHO) estimates that millions of people worldwide suffer from waterborne diseases each year due to inadequate water quality. Water quality is essential for the health of ecosystems, wildlife, and human populations. Contaminated water sources pose significant risks to public health, as waterborne diseases are a leading cause of illness and death worldwide. These diseases, often resulting from the consumption of water polluted with pathogens, chemicals, and heavy metals, impose a substantial burden on society, particularly in vulnerable and underserved communities. Moreover, beyond the immediate human health concerns, compromised water quality also leads to environmental degradation, adversely affecting aquatic ecosystems, biodiversity, and the overall sustainability of natural resources. Consequently, the importance of monitoring and maintaining water quality cannot be overstated. In the big picture, water quality analysis and evaluation techniques have substantially improved the efficiency of water pollution control [1]. To date, many methods have been developed to monitor and assess water quality worldwide, such as the multivariate statistical method [2], fuzzy inference [3], and the water quality index (WQI) [4].

Various pollutants and contaminants can compromise the quality of water sources. These include heavy metals like lead, cadmium, and mercury, pathogenic microorganisms such as bacteria and viruses, and chemical compounds like nitrates and arsenic. The presence of these contaminants in drinking water can have dire consequences for public health, causing diseases such as cholera, dysentery, and lead poisoning. Detecting and classifying water as safe or unsafe as a complex and multifaceted challenge. Although highly accurate, traditional laboratory-based methods for water quality assessment are often time-consuming, resource-intensive, and not conducive to real-time monitoring. Therefore, there is a pressing need for innovative approaches that can provide timely and reliable assessments of water quality.

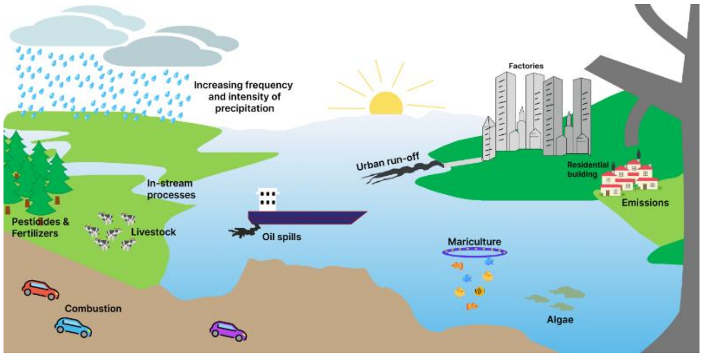


Fig. 1. Contributors towards poor quality of water.

Fig. 1 illustrates the large number of contributors that influence the quality of water. These contributors are entry points for various elements and chemicals that can significantly affect water quality. Just as depicted in the figure, through these various sources and places, a multitude of chemicals are introduced, ultimately influencing the overall quality of water. The objectives for the research are as follows:

- Investigate the utility of machine learning models for water quality classification using a comprehensive set of water quality indicators [5].

- Evaluate and compare the performance of various machine learning algorithms in classifying water sources as safe or unsafe for human consumption.

- Identify critical water quality indicators and features that strongly influence water quality classification, offering insights for targeted monitoring and intervention strategies [6].

- Bridge the gap between traditional water quality assessment methods and emerging technologies to enhance the efficiency and timeliness of water quality monitoring.

- Acknowledge the limitations of considering all water quality parameters due to cost and technical challenges and address the need for more data-driven approaches using machine learning advancements.

The significance of this research extends beyond the confines of academia. It holds practical and societal implications that resonate with the global need for clean and safe drinking water. By developing accurate machine learning models for water quality classification, we contribute to the broader efforts to ensure access to safe and clean drinking water for all. Furthermore, the insights gained from this research have the potential to inform targeted water quality monitoring strategies, enabling more efficient resource allocation and rapid intervention when unsafe water sources are detected. Ultimately, the research aligns with the United Nations Sustainable Development Goals, particularly Goal 6: "Ensure availability and sustainable management of water and sanitation for all", by enhancing our capacity to safeguard water resources and protect human health.

The paper is divided into the following sections: Section II delves into an extensive literature survey, identifying existing research gaps and showcasing innovative approaches in the field. Section III intricately details the system model and architecture, comprising a thorough dataset overview, data preprocessing techniques, and model evaluation methods. The presentation of results is encapsulated in Section IV, featuring performance metric graphs, a confusion matrix, and visual representations of feature importance. Section V navigates through in-depth discussions on practical implications and outlines potential avenues for future research. Finally, Section VI encapsulates the conclusions, summarizing key findings.

II. LITERATURE REVIEW

Li, Z., Liu, H., Zhang, C., & Fu, G. [7] introduced a real-time water quality prediction method for distribution networks using Graph Neural Networks. Addressing sparse monitoring data challenges, the approach underscores GNNs' effectiveness in capturing complex relationships.

Garabaghi, F. [8] employed the AdaBoost ensemble method to classify water sources as safe or unsafe for human consumption based on various water quality indicators. By combining these indicators, their model demonstrated the potential of machine learning to ensure access to safe drinking water. This study contributes to public health and environmental protection efforts.

Li, L. [9] tackled the vital issue of model interpretability in water quality assessment. They introduced a method that combined Random Forest with Shapley values to provide insights into the features contributing to water quality predictions. This research emphasized the importance of model transparency and interpretability in building trust in automated water quality assessment systems.

Cruz, R. [10] explored the application of Machine Learning for predicting harmful algal blooms. Their study utilized historical data on water quality parameters and algal bloom occurrences, demonstrating the potential of data-driven approaches in addressing ecological threats and water safety concerns.

Yan, J. [11] introduced a hybrid model that combined Neural Networks and Principal Component Analysis (PCA) for water quality prediction. Their research emphasized the importance of feature reduction and dimensionality reduction techniques in enhancing the efficiency and effectiveness of predictive models.

Ighalo, J. O. [12] proposed a novel approach that integrated Internet of Things (IoT) technology with machine learning for real-time water quality monitoring. Their study focused on sensor networks and data analytics, enabling proactive responses to water quality deviations.

Barzegar, R. [13] employed Extreme Learning Machines (ELM) for water quality prediction. Their research highlighted the speed and efficiency of ELM in handling large datasets, making it a valuable tool for real-time monitoring and forecasting.

Mosavi, A. [14] researched water quality assessment using ensemble models. By combining Random Forest, Gradient Boosting, and AdaBoost, their approach improved the robustness and accuracy of predictions, addressing the need for reliable water safety assessments.

Li, L. [15] explored the application of Recurrent Neural Networks (RNNs) for predicting temporal water quality variations. Their study emphasized the importance of considering historical data and dynamic patterns in water quality assessment.

Kadinski, L. [16] proposed a data-driven approach to identify contamination sources in water distribution systems. By integrating machine learning and network analysis, their research contributed to the early detection and management of waterborne risks.

Haghiabi, A. H. [17] introduced a framework for water quality prediction using multiple machine learning models. Their approach combined Support Vector Machines, Decision Trees, and K-nearest neighbors to improve predictive accuracy and model robustness.

Chakravarthy [18] introduced a method focusing on water quality prediction, employing SoftMax-ELM optimized with the Adaptive Crow-Search Algorithm. This innovative technique aims to enhance accuracy in water quality predictions by optimizing the SoftMax-ELM model.

Dogo, E. M. [19] explored using unsupervised learning for water quality anomaly detection. Their research applied Self-Organizing Maps (SOM) to identify deviations from normal water quality conditions, enhancing the early detection of water contamination incidents.

Solanki, A. [20] laid the foundation for machine learning models in water quality assessment. Their early work paved the way for subsequent research by highlighting the potential of data-driven approaches in this domain. This study marks the inception of applying machine learning to water quality analysis.

Chang, N. B. [21] explored integrating remote sensing data with machine learning techniques to assess water quality in large water bodies. Their approach showcased the scalability of machine learning in monitoring vast aquatic environments, with implications for environmental conservation and management.

Wu, J. [22] delved into using Long Short-Term Memory (LSTM) neural networks for time-series-based water quality prediction. Their study focused on predicting temporal variations in water quality indicators, aiding in forecasting changes over time. This research contributes to a better understanding of the dynamic nature of water quality.

Moayedi, H. [23] developed a hybrid model that combined machine learning and physical models for water quality assessment. This integrated approach improved prediction accuracy by considering both data-driven and mechanistic aspects, bridging the gap between empirical and theoretical approaches in water quality research.

Yan, K. [24] introduced a novel feature engineering technique called Recursive Feature Extraction (RFE) for water quality data. This method improved model performance by selecting the most relevant features for prediction, enhancing the efficiency and effectiveness of water quality assessment models.

Ahmed, M. [25] explored the application of Deep Belief Networks (DBNs) for water quality monitoring. Their study highlighted the potential of deep learning techniques in capturing complex patterns in water quality data, paving the way for advanced modeling approaches in the field.

Liu, S. [26] investigated the use of evolutionary algorithms in optimizing machine learning models for water quality prediction. Their research emphasized the importance of model tuning and parameter optimization for improved prediction accuracy, contributing to more reliable water quality assessments.

Iqbal, K. [27] applied clustering techniques to segment water quality data into distinct groups. This unsupervised learning approach facilitated.

Identifying common patterns and anomalies in water quality profiles offers insights for targeted monitoring and intervention strategies.

Table I provides the research gaps in the previous approaches for tackling the issues with water quality.

A. Research Gaps

- Data-Driven Insights for Targeted Monitoring: The paper fills a research gap by providing insights into data-driven approaches that help identify critical water quality indicators and features.
- Integration of Various Machine Learning Models: The paper addresses the gap in research related to integrating and comparing multiple machine learning models for water quality assessment, providing insights into the performance of different algorithms.
- Interpretability and Transparency: The paper bridges the gap by addressing the need for model interpretability and transparency in the context of water quality assessment

In Fig. 2, we visually represent the diverse contributions from various fields that have shaped and influenced our research. This figure illustrates the multidisciplinary nature of the research endeavor and the collaborative efforts of experts from different domains.

TABLE I. RESEARCH GAPS

Ref No	Author	Proposed Method	Limitations
3	Garabaghi, F.	AdaBoost for classifying water sources as safe or unsafe	Limited discussion of model performance and real-world application challenges.
17	Chang, N. B.	Integration of remote sensing with ML for assessing water quality	Challenges in remote sensing data quality and applicability to different ecosystems.
8	Ighalo, J. O.	IoT and ML for real-time water quality monitoring	Potential issues with sensor network deployment, data quality, and security.
10	Mosavi, A.	Ensemble models combining RF, GB, and AdaBoost for prediction	Complexity in model interpretation and computational resources required.
23	Iqbal, K.	Clustering to segment water quality data into distinct groups	Dependency on the quality of input data and the choice of clustering algorithm.

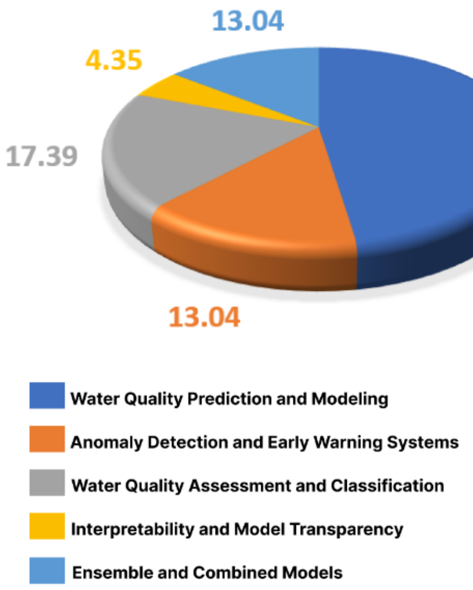


Fig. 2. Contributions for the research.

III. System Model and Architecture

A. Dataset Overview

The dataset was obtained from Kaggle, a well-known platform for sharing and exploring datasets. The dataset contains information related to water quality parameters and attributes for classifying water sources as safe or unsafe for human consumption. The dataset comprises 21 columns and 8000 rows, providing substantial data for robust analysis and modeling.

Below are the dataset values and their significance in deteriorating or enhancing water quality.

1) *Aluminum:* Measures the concentration of aluminum in the water.

2) *Ammonia:* Indicates the ammonia level in the water.

3) *Arsenic:* Reflects the concentration of arsenic in the water.

4) *Barium:* Represents the amount of barium in the water.

5) *Cadmium:* Measures the cadmium content in the water.

6) *Chloramine:* Indicates the chloramine level in the water.

7) *Chromium:* Reflects the concentration of chromium in the water.

8) *Copper:* Measures the copper content in the water.

9) *Fluoride:* Represents the fluoride level in the water.

10) *Bacteria:* Reflects the presence or absence of bacteria in the water.

11) *Viruses:* Indicates the presence or absence of viruses in the water.

12) *Lead:* Measures the lead content in the water.

13) *Nitrates:* Reflects the nitrate level in the water.

14) *Nitrites:* Indicates the nitrite level in the water.

15) *Mercury:* Measures the mercury content in the water.

16) *Perchlorate:* Represents the amount of perchlorate in the water.

17) *Radium:* Reflects the concentration of radium in the water.

18) *Selenium:* Measures the selenium content in the water.

19) *Silver:* Indicates the silver level in the water.

20) *Uranium:* Reflects the concentration of uranium in the water.

21) *Is_safe:* The class attribute, where '0' represents not safe and '1' represents safe water sources.

B. Data Preprocessing and Feature Engineering

Given the diverse range of water quality indicators, such as aluminum, ammonia, arsenic, and others, the data preprocessing phase involved several key steps. We addressed missing values by mean imputation to prevent gaps in the dataset, normalized the values of these indicators, and encoded categorical attributes. Eq. (1) is used to handle missing values.

$$x_i = \frac{1}{N}\sum_{j=1}^{N} x_j \qquad (1)$$

Moreover, the most critical aspect of feature engineering was the transformation of raw indicator values into binary representations. This transformation enabled us to create the "is_safe" feature, serving as the class attribute for classification and ultimately facilitating accurate water quality assessment. To facilitate the modeling process, we employed the StandardScaler function to scale the features into a common range. This normalization was vital for ensuring that each indicator contributed to the classification process on an equal footing. For a given feature X, the standardization using StandardScaler transforms it into a new feature X', given in Eq. (2). This transformation ensures that the standardized feature has a mean of 0 and a standard deviation of 1, which is important for many machine learning algorithms, especially those sensitive to the features' scale.

$$X' = \frac{X-\mu}{\sigma} \qquad (2)$$

where:

X' is the standardized feature.

X is the original feature.

μ is the mean(average) of the feature X.

σ is the standard deviation of the feature X.

C. Model Evaluation

Evaluating the performance of machine learning models for water quality classification is a critical aspect of our research. To gauge the effectiveness of these models, we considered multiple performance metrics. Accuracy was an essential metric that measures the proportion of correctly classified instances. We also examined the F1 score, which balances precision and recall, ensuring that our models can efficiently identify both safe and unsafe water sources. Furthermore, precision and recall were critical for understanding the model's ability to minimize false positives and false negatives, respectively. The ROC AUC score assessed the models' ability to distinguish between safe and unsafe sources. By employing these metrics, our research ensures rigorous evaluation and validation of the water quality classification models.

Algorithm 1 shows how to calculate the accuracy of a machine-learning model. As more accurate model outcomes result in better decisions, it is important to identify which model would work best for a given dataset.

Algorithm 1: To Calculate Accuracy for Machine Learning Model

```
Input:
- Trained machine learning model (ML_model)
- Test data (X_test) with corresponding true labels (y_true)
Output:
- Accuracy of the machine learning model
1. Initialize a variable 'correct_predictions' to 0.
2. Initialize a variable 'total_predictions' to 0.
3. For each data point (x) and true label (y_true) in the test data (X_test, y_true):
   a. Use the trained machine learning model (ML_model) to make a prediction (y_pred) for x.
   b. Increment 'total_predictions' by 1.
   c. If the model's prediction (y_pred) matches the true label (y_true):
      - Increment 'correct_predictions' by 1.
4. Calculate the accuracy as follows:
   - Accuracy = (correct_predictions / total_predictions) * 100
5. Output the accuracy value as the result.
End
```

Algorithm 2 shows how the calculation for ROC AUC score is performed and how the plot for the ROC curve is designed. The ROC AUC score tells us how efficient the model is. The higher the AUC, the better the model's performance at distinguishing between the positive and negative classes. An AUC score of 1 means the classifier can perfectly distinguish between all the Positive and the Negative class points.

Algorithm 2: To Calculate ROC AUC Score and Plot ROC Curve

```
Input:
- Trained machine learning model (ML_model)
- Test data (X_test) with corresponding true binary labels (y_true)
Output:
- ROC AUC Score
- ROC Curve Plot
1. Use the trained machine learning model (ML_model) to predict probabilities for each data point in the test data (X_test).
2. Calculate the ROC AUC score using the predicted probabilities and true labels:
   - ROC_AUC_Score = roc_auc_score(y_true, predicted_probabilities)
3. Compute the ROC curve by varying the decision threshold:
   - FPR (False Positive Rate), TPR (True Positive Rate), thresholds = roc_curve(y_true, predicted_probabilities)
4. Plot the ROC curve:
   - Plot FPR is on the x-axis, and TPR is on the y-axis.
   - Add a diagonal line representing a random classifier (FPR = TPR) for reference.
   - Label the curve and the diagonal line accordingly.
   - Add a legend to the plot.
5. Output the ROC AUC Score and the ROC Curve Plot.
End
```

C. Model Evaluation Equations for Machine Learning Algorithms

The selection of appropriate models plays a crucial role in the success of any research endeavor, as different models employ distinct algorithms and mathematical techniques to model the underlying data. This subsection presents an overview of the machine learning models employed in this research, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). Each model's Equation is discussed in Eq. (3) to Eq. (7), elucidating the mathematical foundation upon which they operate, allowing for a comprehensive understanding of their implementation in the context of this research.

Logistic Regression:

$$P(y = 1|x) = 1 / (1 + exp(-z) \tag{3}$$

where:

P(y=1|x) is the likelihood that the positive class will exist.

z is the linear combination of the input features and their corresponding coefficients.

Decision Tree:

$$Prediction = Tree(x) \tag{4}$$

where:

Tree(x) represents the traversal of the decision tree to assign the class label to the instance x based on the

Decision Tree learned rules.is a tree-based classifier that splits the feature space based on a set of rules.

Random Forest:

$$Prediction = Average(Tree1(x), Tree2(x), \ldots, TreeN(x) \tag{5}$$

where:

Tree(x) represents the traversal of the decision tree.

An ensemble technique called Random Forest blends various decision trees to produce estimations.

The prediction in a Random Forest is obtained by averaging the predictions of individual decision trees.

Support Vector Machines (SVM):

$$Prediction = sign(w^\wedge T * c + e) \tag{6}$$

where:

Prediction is the predicted class label.

w is the weight vector.

c are the input features.

e is the bias term.

The sign function assigns the class label based on the sign of the linear combination.

Support Vector Machines are binary classifiers that aim to find the hyperplane that maximizes the margin between two classes.

KNeighborsClassifier:

$$dist(x, z)\ =\ (\Sigma\_\{r = 1\}\^{}d\ |x\_r\ -\ z\_r|\^{}p)\^{}(1/p) \qquad (7)$$

where:

dist(x,z) represents the distance between the two points x and z.

Σ_{r=1}^d sums all the features of the data.

|x_r - z_r|^p calculates the absolute difference in each dimension.

D. Architecture

Fig. 3 provides an overview of how data processing works in our paper by showing each step involved in the process from input to output as well as illustrating how information flows between these steps.

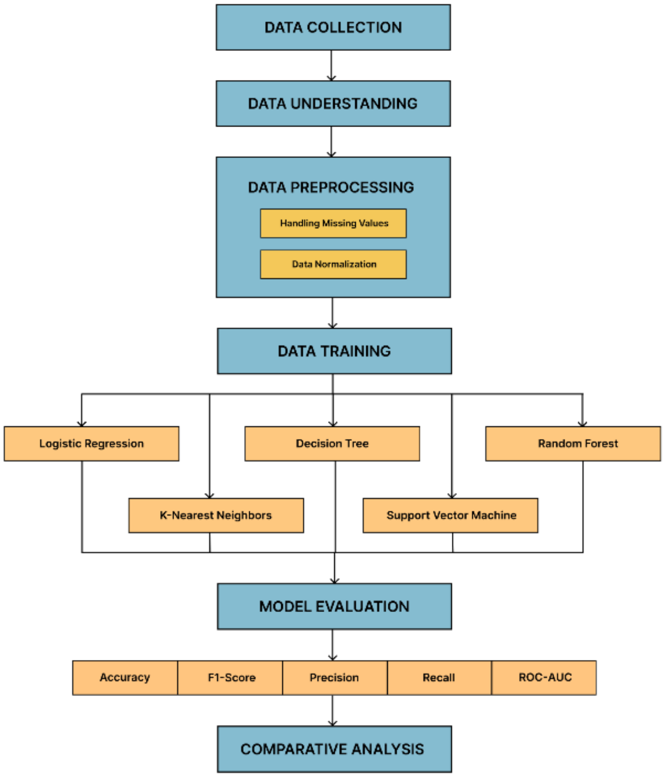


Fig. 3. Architecture diagram.

IV. Results

Our research, produced significant findings that hold practical implications for ensuring safe drinking water. Fig. 4 highlights the Accuracy of our machine learning models on the test set. Notably, the Random Forest model achieved an accuracy of 95.08% reflecting its robustness in classifying water sources as safe or unsafe. The Decision Tree model closely followed, demonstrating an accuracy of 94.62%. Logistic Regression, K-Nearest Neighbors, and Support Vector Machine also provided competitive results. Fig. 5 highlights the F1-Score for the various models. The Random Forest model achieved a score of 94.69%, followed by Decision Tree model with a score of 94.63%. Fig. 6 highlights the Precision for the models. Random Forest model achieved a score of 90.48%, which is followed by Support Vector Machine model with a score of 86.58 %. Fig. 7 highlights the Recall of the various models used. Random Forest model gave the best score of 93.10%, which is followed by Support Vector Machine model with a score of 87.60%.

Fig. 8 illustrates the ROC-AUC curve, which depicts the performance of different machine learning models in classifying water quality ranging from a score of 0 to 1. Random Forest outperforms other models with the highest AUC score of 0.91, showcasing its superior ability to discriminate between various water quality levels.

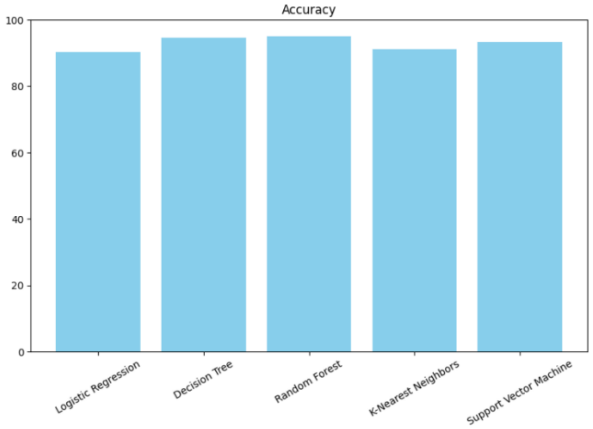


Fig. 4. Model Comparison in terms of accuracy.

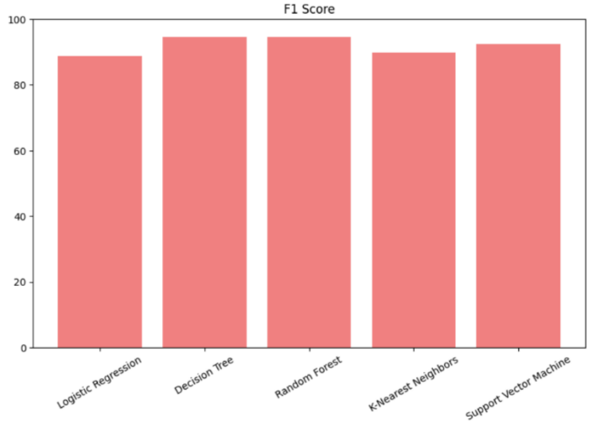


Fig. 5. Model comparison in terms of F1-score.

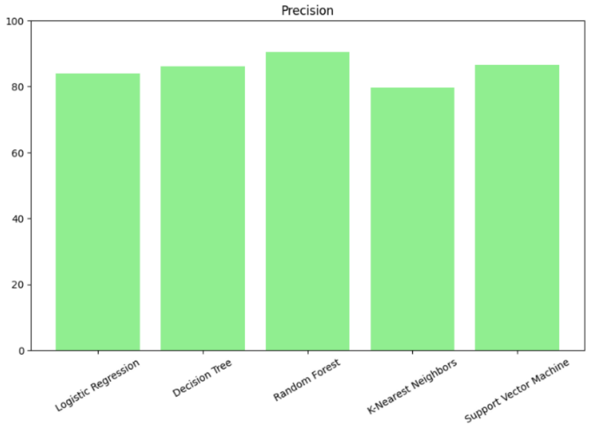


Fig. 6. Model Comparison in terms of precision.

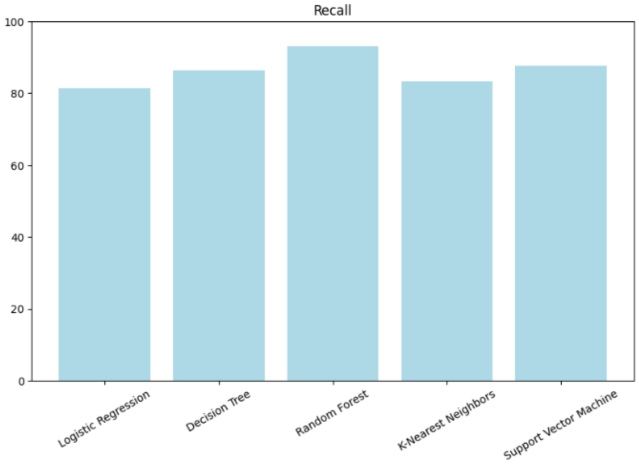


Fig. 7. Model comparison in terms of recall.

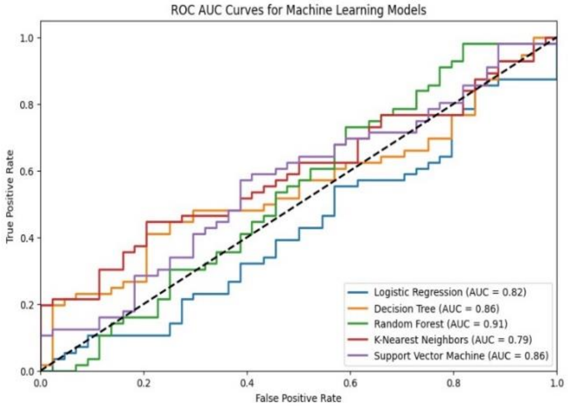


Fig. 8. ROC AUC score.

Feature importance is typically calculated based on how often a feature is selected to split nodes in decision trees within the ensemble and how much the feature contributes to reducing impurity in those splits. Feature importance using a Random Forest model can help identify which water quality parameters have the greatest impact on predicting water quality outcomes. This knowledge is essential for understanding the most influential factors affecting water quality, enabling better decision-making. Fig. 9 illustrates the results.

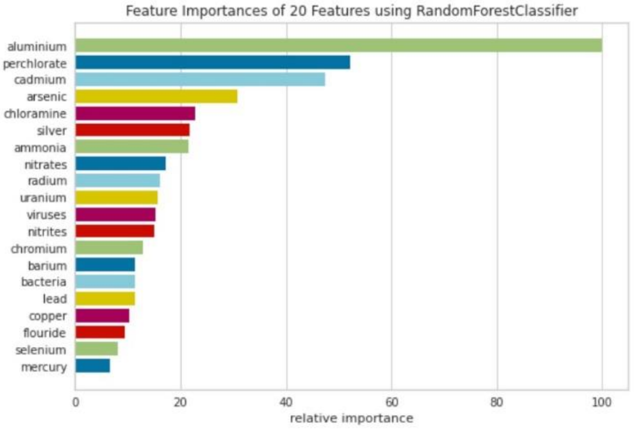


Fig. 9. Feature importance.

Table II presents the importance of the feature obtained from a random forest model. The "Feature" column lists the input variables, while the "Importance" column quantifies the significance of each feature in the model's predictions. These importance scores, ranging from 0 to 1, reveal the relative influence of each variable in detecting the water quality.

TABLE II. IMPORTANCE OF A SPECIFIC FEATURE

Feature	Importance
aluminum	0.214372
perchlorate	0.119331
cadmium	0.113347
arsenic	0.065622
ammonia	0.046725
chloramine	0.046548
Silver	0.045136
nitrates	0.037736
nitrites	0.033658
uranium	0.033620
radium	0.032812
viruses	0.031122
chromium	0.029673
barium	0.028580
bacteria	0.026482
lead	0.023878
copper	0.023279
fluoride	0.019532
selenium	0.016568
Mercury	0.011981

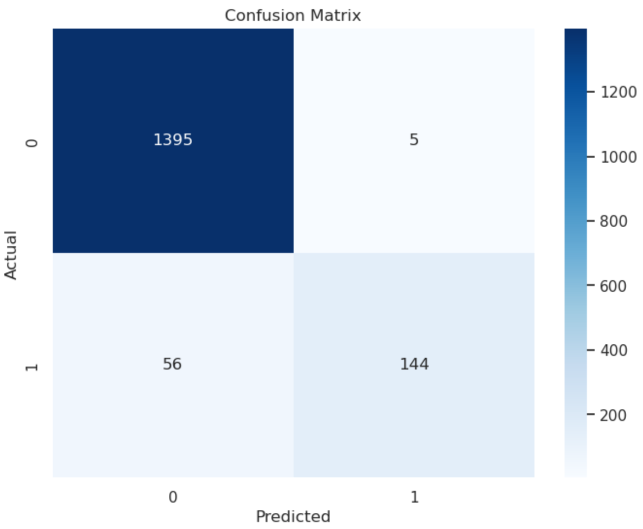


Fig. 10. Confusion matrix.

The confusion matrix illustrated in Fig. 10 visually represents the classification performance of the random forest model. It offers a detailed breakdown of true positives, true negatives, false positives, and false negatives, providing insights into the model's accuracy and error patterns.

The heatmap in Fig. 11 represents the relationships between the different water quality features in the dataset, to help identify which parameters are strongly correlated (positively or negatively) and which are not significantly related.

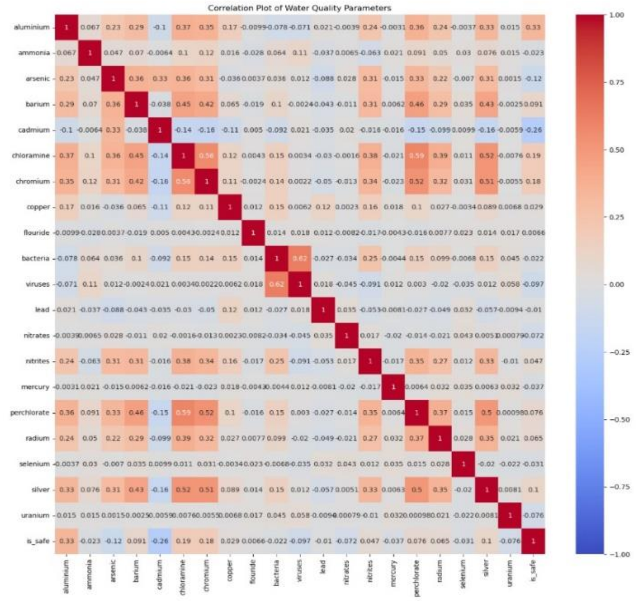


Fig. 11. Heatmap between water quality features.

TABLE III. SUMMARY OF RESULTS

Model	Accuracy	F1-Score	Precision	Recall	ROC-AUC
Logistic Regression	90.25	88.73	83.91	81.37	0.82
Decision Tree	94.62	94.63	86.10	86.38	0.86
Random Forest	95.08	94.69	90.48	93.10	0.91
K-Nearest Neighbors	91.08	89.86	79.66	83.27	0.79
Support Vector Machine	93.25	92.36	86.58	87.60	0.86

Table III provides a comparative analysis of all the models used and their performance measures.

V. DISCUSSIONS

A. *Interpretation of Results*

The interpretation of our research results reveals a comprehensive understanding of their implications for water quality assessment. Our machine learning models, especially the Random Forest and Decision Tree, have proven their capability to effectively classify water sources as safe or unsafe. This has significant practical implications, particularly in the context of providing safe drinking water. The high accuracy and F1 scores signify the models' ability to minimize false positives and negatives, an essential characteristic when dealing with public health issues related to water quality. These results demonstrate the potential for real-time water quality monitoring, allowing for the swift detection of anomalies and timely intervention.

B. *Feature Importance*

Analyzing the importance of features, as highlighted in our feature importance plots, provides valuable insights into the significant indicators influencing water quality classification. This knowledge empowers decision-makers and water quality management authorities to prioritize interventions. By identifying which indicators have the most substantial impact, targeted actions can be taken to ensure water safety. The binary representation of these features makes it easy to understand and act upon the results. Furthermore, feature importance analysis complements traditional laboratory-based methods by providing a data-driven approach to identifying critical water quality indicators.

C. *Practical Implications*

The practical implications of our research are substantial and extend far beyond the scope of our findings. Our models have showcased their potential for real-time water quality monitoring, which can be a game-changer in ensuring the safety of drinking water. The ability to rapidly detect water sources that pose health risks can transform public health management. This research is especially relevant in regions where water quality can fluctuate significantly, potentially impacting the health of communities. By harnessing machine learning models, authorities and stakeholders can efficiently monitor and manage water quality, taking timely actions to address concerns.

D. Limitations

Our research, while promising, relies on historical data, which may not fully capture evolving water quality dynamics. Additionally, our approach assumes fixed threshold values for safety, which may not be universally applicable across different regions and water sources. It is essential to recognize that water quality can vary significantly due to geographical and environmental factors. Future research should aim to address these limitations by considering real-time data and accounting for regional variations. In conclusion, our research has provided valuable insights into the potential of machine learning models for water quality assessment. The interpretability of results, the identification of significant features, and the practical implications of our findings underscore the significance of automated systems in ensuring safe drinking water. The limitations highlighted here serve as a roadmap for future research endeavors to continually improve water quality management and public health worldwide.

E. Future Scope

The success of our current research opens up several promising avenues for future investigations in the field of water quality assessment and management. Some areas where further research and development can make significant contributions are:

1) Real-time data integration: Future research should focus on integrating real-time data sources into the machine learning models. By leveraging the power of continuous data streams from various sensors and sources, we can create more adaptive and responsive models that can detect water quality anomalies as they happen. This would enhance the timeliness and accuracy of intervention strategies.

2) Advanced sensor technology: Researchers can explore the development of cutting-edge sensor technologies that can provide more granular data on water quality parameters. This could involve using nano-sensors, microfluidic devices, and remote sensing technologies to measure a wide range of chemical, biological, and physical indicators in real time.

3) Deep learning and neural networks: Investigate the application of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), for analyzing complex, high-dimensional water quality data.

4) Explainable AI (XAI): Develop explainable AI techniques to enhance the interpretability of machine learning models. This is crucial for gaining the trust of stakeholders and decision-makers in the water quality management process. Methods such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) can be employed to provide insights into model predictions.

5) Edge computing and real-time processing: Leverage edge computing and real-time data processing to reduce anomaly detection and response latency. Edge devices equipped with machine learning capabilities can make rapid decisions on data streams, enabling timely intervention.

6) Blockchain for data verification: Use blockchain technology to enhance data integrity and trustworthiness. Blockchain can be employed to securely record and verify water quality data, ensuring its accuracy and preventing tampering.

VI. Conclusion

The research highlights the efficacy of machine learning models in classifying water quality, offering significant practical implications for ensuring safe drinking water. The Random Forest model stood out as the top performer, achieving an accuracy of 95.08% and an F1-score of 94.69. Its precision of 90.48% and recall of 93.10% underscore its ability to identify safe water sources while minimizing false alarms accurately. The ROC-AUC curve further emphasizes the Random Forest's superiority, with the highest AUC of 0.91, signifying its reliability in discriminating between water quality levels. Feature importance analysis using the Random Forest model unveiled crucial insights into the most influential factors affecting water quality outcomes, providing valuable knowledge for decision-making in water quality management.

In summary, this study demonstrates that machine learning, particularly the Random Forest algorithm, is a powerful tool for classifying water quality with high accuracy. As far as the practical implications are considered, this research can be applied to the regions where water quality can fluctuate significantly. By harnessing machine learning models, authorities and stakeholders can efficiently monitor and manage water quality, taking timely actions to address concerns. These findings can inform policies and strategies to ensure clean and safe water sources, ultimately enhancing environmental and public health.

References

[1] Alam, R., Ahmeahd, Z., Seefat, S. M., & Nahin, K. T. K. (2021). Assessment of surface water quality around a landfill using multivariate statistical method, Sylhet, Bangladesh. Environmental Nanotechnology, Monitoring & Management, 15, 100422.

[2] Oladipo, J. O., Akinwumiju, A. S., Aboyeji, O. S., & Adelodun, A. A. (2021). Comparison between fuzzy logic and water quality index methods: A case of water quality assessment in Ikare community, Southwestern Nigeria. Environmental Challenges, 3, 100038.

[3] Wang, J., Fu, Z., Qiao, H., & Liu, F. (2019). Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China. Science of the Total Environment, 650, 1392-1402.

[4] dos Santos Simoes, F., Moreira, A. B., Bisinoti, M. C., Gimenez, S. M. N., & Yabe, M. J. S. (2008). Water quality index as a simple indicator of aquaculture effects on aquatic bodies. Ecological indicators, 8(5), 476-484.

[5] Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., ... & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. Eco-Environment & Health.

[6] Jalal, D., & Ezzedine, T. (2020, June). Decision tree and support vector machine for anomaly detection in water distribution networks. In 2020 International Wireless Communications and Mobile Computing (IWCMC) (pp. 1320-1323). IEEE.

[7] Li, Z., Liu, H., Zhang, C., & Fu, G. (2023). Real-time water quality prediction in water distribution networks using graph neural networks with sparse monitoring data. Water Research, 121018.

[8] Garabaghi, F. H., Benzer, S., & Benzer, R. (2022). Performance evaluation of machine learning models with ensemble learning approach in classification of water quality indices based on different subset of features.

[9] Li, L., Qiao, J., Yu, G., Wang, L., Li, H. Y., Liao, C., & Zhu, Z. (2022). Interpretable tree-based ensemble model for predicting beach water quality. Water Research, 211, 118078.

[10] Cruz, R. C., Reis Costa, P., Vinga, S., Krippahl, L., & Lopes, M. B. (2021). A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination. Journal of Marine Science and Engineering, 9(3), 283.

[11] Yan, J., Liu, J., Yu, Y., & Xu, H. (2021). Water quality prediction in the luan river based on 1-drcnn and bigru hybrid neural network model. Water, 13(9), 1273.

[12] Ighalo, J. O., Adeniyi, A. G., & Marques, G. (2021). Internet of things for water quality monitoring and assessment: a comprehensive review. Artificial intelligence for sustainable development: theory, practice and future applications, 245-259.

[13] Barzegar, R., Asghari Moghaddam, A., Adamowski, J., & Ozga-Zielinski, B. (2018). Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stochastic environmental research and risk assessment, 32, 799-813.

[14] Mosavi, A., Hosseini, F. S., Choubin, B., Abdolshahnejad, M., Gharechaee, H., Lahijanzadeh, A., & Dineva, A. A. (2020). Susceptibility prediction of groundwater hardness using ensemble machine learning models. Water, 12(10), 2770.

[15] Li, L., Jiang, P., Xu, H., Lin, G., Guo, D., & Wu, H. (2019). Water quality prediction based on recurrent neural network and improved evidence theory: a case study of Qiantang River, China. Environmental Science and Pollution Research, 26, 19879-19896.

[16] Kadinski, L., Salcedo, C., Boccelli, D. L., Berglund, E., & Ostfeld, A. (2022). A hybrid data-driven-agent-based modelling framework for water distribution systems contamination response during COVID-19. Water, 14(7), 1088.

[17] Haghiabi, A. H., Nasrolahi, A. H., & Parsaie, A. (2018). Water quality prediction using machine learning methods. Water Quality Research Journal, 53(1), 3-13.

[18] Chakravarthy, S. S., Bharanidharan, N., kumar Venkatesan, V., Abbas, M., Rajaguru, H., Mahesh, T. R., & Venkatesan, K. (2023). Prediction of Water Quality using SoftMax-ELM optimized using Adaptive Crow-Search Algorithm. IEEE Access.

[19] Dogo, E. M., Nwulu, N. I., Twala, B., & Aigbavboa, C. (2019). A survey of machine learning methods applied to anomaly detection on drinking-water quality data. Urban Water Journal, 16(3), 235-248.

[20] Solanki, A., Agrawal, H., & Khare, K. (2015). Predictive analysis of water quality parameters using deep learning. International Journal of Computer Applications, 125(9), 0975-8887.

[21] Chang, N. B., Bai, K., & Chen, C. F. (2017). Integrating multisensor satellite data merging and image reconstruction in support of machine learning for better water quality management. Journal of environmental management, 201, 227-240.

[22] Wu, J., & Wang, Z. (2022). A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. Water, 14(4), 610.

[23] Moayedi, H., Salari, M., Dehrashid, A. A., & Le, B. N. (2023). Groundwater quality evaluation using hybrid model of the multi-layer perceptron combined with neural-evolutionary regression techniques: case study of Shiraz plain. Stochastic Environmental Research and Risk Assessment, 1-16.

[24] Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sensors and Actuators B: Chemical, 212, 353-363.

[25] Ahmed, M., Mumtaz, R., Anwar, Z., Shaukat, A., Arif, O., & Shafait, F. (2022). A multi–step approach for optically active and inactive water quality parameter estimation using deep learning and remote sensing. Water, 14(13), 2112.

[26] Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., & Wei, Y. (2013). A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Mathematical and Computer Modelling, 58(3-4), 458-465.

[27] Iqbal, K., Ahmad, S., & Dutta, V. (2019). Pollution mapping in the urban segment of a tropical river: is water quality index (WQI) enough for a nutrient-polluted river?. Applied Water Science, 9(8), 1-16.

Data Mining Application Forecast of Business Trends of Electronic Products

Kheo Chau Mui, Nhon Nguyen Thien
Information Technology Department, FPT University, Cantho city, Vietnam

***Abstract*—Sales forecasting is a pressing concern for companies amid rising consumer demand and intensifying competition, compounded by declining sales due to growing socio-economic challenges. Currently, many companies are having difficulty selling products due to a lack of management systems. To assist that, data mining techniques are introduced but it is difficult to evaluate the data and it is practically impossible to accurately forecast large amounts of data. However, data mining remains an important management tool that supports early decisions to increase profits, innovate business trends and improve sales by generating intelligence from the company's data resources. In this article, the research object chosen is the data of a nationwide electronics company. Their sales volume data for consumer electronics was used and applied to this study. The study used a "clustering" algorithm to group data based on the unique characteristics of each product, region, season, and time to estimate the amount of goods sold in the past, thereby predicting the amount of goods that will be exported. Password is sold in the following years and look for market trends. For each group, the results obtained with k = 3 show that the number of elements in each cluster is 771422, 11874, and 312, respectively. Combined with the "regression tree" algorithm for cluster partitioning and using the protocol Evaluate MSE and RMSE to evaluate the accuracy of the model, a result of 43065.66 Sales forecasting results show that the model's accuracy is close to realistic accuracy and depends on seasonal factors that are really important to some people. Based on the above results, the business's marketing campaigns and strategies will be deployed and achieve high results.**

Keywords—Data mining; sales forecasting; clusters; regression tree; RMSE; MSE; k-prototypes

I. Introduction

Electronic gadgets have steadily become important goods to assist people's life in the contemporary era of smart technology. This generates a lucrative profit for firms dealing in electronic items; as rivalry between businesses grows, they must always come up with new business advantages to compete [40], [41]. In the electronics business, you must compete and survive. These benefits must be based on the quantity of data gathered in the past and present from internal operations, product supply procedures, market trends and the business environment, and consumer preferences, to analyse patterns, estimate future sales, and establish a change management approach that brings efficiency and quality to corporate management while saving operational costs and expenses. Costs of storing are reduced, and profits are boosted [1], [5]. However, achieving high company efficiency through data mining, analysis, and trend prediction is a challenging task for firms. As a result, a system is required for firms to efficiently explore, analyse, process, and anticipate new business trends [27], [28].

Previous research has offered an overview of contemporary obstacles in sales forecasting as well as difficulties in trading electronic products, such as:

- Product life cycle is becoming more shorter.
- Consumer demand has increased and become more diverse.
- The average industrial land rental price rises by 5-8% every year.

To be very lucrative, a corporation must properly estimate the output of commodities, at the right moment of demand, and restrict inventory. Businesses will have challenges without the tools of information processing and analysis to assist anticipate the next business condition if they have a large and diversified data source. It is important to develop a "Data mining application to forecast business trends for electronic products" based on relevant research and existing business practices. This programme will assist businesses in selecting a strategy to anticipate the optimal output of items for their firm based on each area, industry, and seasonal features.

II. Related Work

For many years, several methodologies have been utilised in sales forecasting research. Here are some examples of typical studies. Pure Classification (PC) and Hybrid Clustering Classification (HCC) are two data mining techniques suggested by Bhavin Parikh et al. [4]. The results of the tests demonstrate that the HCC model outperforms the PC model in terms of accuracy and performance. In particular, after evaluating 500 samples with Nearest Neighbours = 5, the accuracy attained is 57.62%, outperforming the PC model. Fifit Alfiah et al. [2] investigated association rules in data mining with excellent dependability. Support: 0.1 and support x Confidence: 0.05 are the results, and if the manager simply enters the forecast quantity of 13 goods, the output forecast quantity is support: 0.2, support x confidence: 0.1, and prediction percentage: 0.15. The quantity value is enhanced by 15% from the initial value based on the outcomes and association rules in data mining. A paper in [3] uses the K-means algorithm to divide data into three separate clusters based on product type and sold quantity, namely Dead-Stock (inventory), Slow-Move (sold products), and rapid-Move (rapid sale). Next, employ the MFP: Most Frequent Pattern algorithm to identify frequent item attribute patterns in each product category while also providing sales trends in a concise format. Lytvynenko Tetiana's [6] research employed two key methodologies: statistical and structural methods, respectively, along with the widely used Decision Tree method to represent the process visually and simply on java. Simplify the main goal of the analysis. They utilised a data set that contained the sales volume of an employee job, the month of sales, and the sales of a business from 2006 to 2009. After sifting through the binary tree, they discovered that April, November, and December had the largest sales (about 23000 units each month), although accounting for 18771 in 2006. 19139 and 15164. Mustapha Ismail et al. [5] conducted research on data mining (DM) for e-commerce, which included three general algorithms: matching, grouping, and prediction. It also discusses some of the advantages of DM for e-commerce businesses, including as item planning, sales forecasting, shopping cart analysis, customer relationship management, and probable market segmentation obtained by the three techniques listed above. Furthermore, this research assesses data mining difficulties such as spider detection, data translation, and making the data model intelligible to business users.

III. Related Methods

To implement the application, we choose to utilise the "K-means" and "K-Prototypes" algorithms to cluster data based on the characteristics of each product, region, and time, in conjunction with the "Regression tree" technique. From [7], [8], [9], [12], [19], [20], [21], [24], [25], [26]. Python is the programming language used in this application. This combination aids in forecasting development patterns, product sales to address the issue of shortages, not keeping up with the seasons, consumer purchasing habits, and undesirable inventory that firms frequently face. Management and operation must entail significant expenditures.

A. K-means

J.A. Hartigan and M. A. Wong of Yale University created the Kmeans algorithm [18]. It is one of the finest algorithms for forming clusters of tiny values from vast amounts of unlabeled data (Label). The Kmeans method is an iterative algorithm that attempts to locate data clusters that are as close as feasible.

The initialization of the number of k groups is the first stage in the Kmeans algorithm [13], [14], [15], [16], [17]. And the starting centre value for each group is chosen at random.

The distance between each element and the centre values of each group is then calculated. Different formulae will be employed to calculate the distance depending on the properties of each data type. Manhattan, Cosine, Minkowski, and Euclidean distance metrics are utilised for numerical data. The Euclidean distance is employed in this research to compute the distance from the centre to the elements:

$$\boldsymbol{d(x_i, x_j)} = \sqrt{\left(|\boldsymbol{x_{i1}} - \boldsymbol{x_{j1|}}|^2 + |\boldsymbol{x_{i2}} - \boldsymbol{x_{j2}}|^2 + \cdots + |\boldsymbol{x_{in}} - \boldsymbol{x_{jn}}|^2\right)}$$

In n-dimensional space, xi=(xi1,xi2,...,xin) and xj=(xj1,xj2,...,xjn) are two data points.

Then, until convergence, perform the following operations:

+ Based on distance, each element is assigned to its nearest centre.

+ Update the centres of K clusters, each centre being the mean value of the elements in its cluster.

B. K-prototypes

Separates the dissimilarity of the combined data into two sections for independent computations. The numerical component employs squared Euclidean distance, whereas the mixed part uses basic mactching distance [37, 38, 39] Because the ratio of the two data types is not the same, the study will alter the parameters in the K-Prototypes method after the calculation to avoid the divergence of the grouping result value. The distance is defined as follows, where m is the number of matches and p is the total number of variables (attributes):

$$d(i,j) = \frac{p-m}{p}$$

The K-Prototypes algorithm is implemented as follows:

Input: Initial data set X and number of clusters k.

Output: k sample objects so that the standard function approaches the minimal value.

1) Create k initial sample objects for X, each of which serves as the representative centre of each cluster.

2) Distribute each X feature to each cluster so that it is closest to the sample object in the cluster, while updating the sample object for each cluster.

3) After all of the objects have been dispersed to the clusters, compare their similarity to the sample objects to see if there is a sample object most similar to it that varies from the other. The current cluster's sample object moves the object under consideration to the cluster corresponding to the sample object nearest to it and simultaneously changes the sample objects for these two clusters.

4) After inspecting all objects, repeat step 3 until no object changes.

C. Elbow Method

The Elbow technique [22] is one approach for determining the ideal number of clusters k while clustering. This approach is based on Thorndike's [23] hypothesis, and Elbow is a visual method. The SSE (Sum of Squared Error) indicator represents this Elbow technique.

The basic idea behind this approach is to compute k values by squaring the distances between the members in each cluster and the cluster centre. The sum of squared errors (SSE) is used for comparison. Repeat the k value and compute the SSE; smaller values suggest that each cluster is more convergent.

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in S_k} ||x_i - C_k||_2^2$$

where: k is the number of clusters, Ck is the kth cluster, and x is the number of cluster elements.

The Elbow approach may be defined in two ways:

1) *Deformation:* It is determined as the mean of the squared distances from the individual cluster centres. The Euclidean distance measure is commonly employed.

2) *Inertia:* The sum of the squares of the sample distances from the nearest cluster centre.

Iterate over the values of k from 1 to 9 and compute the distortion and inertia for each value of k in the specified range shown in Fig. 1.

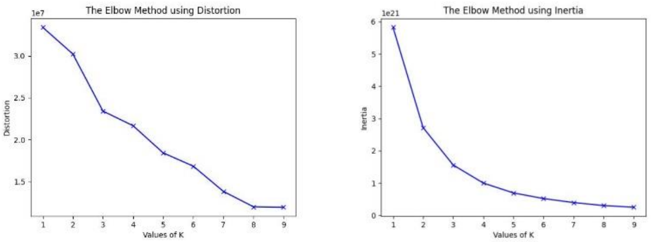


Fig. 1. An illustration of the elbow method.

D. Regression Tree Method

The regression tree approach is used to forecast continuous label values like revenue, profit, product cost, and so on.

An algorithm-based regression tree is used to estimate sales trends ([10]). In terms of current advancements in regression tree approaches, see [29], [30], [31], [32], [34], [35], [36.Regression trees are built using a method known as binary recursive partitioning, which is an iterative procedure that divides data into branches and then further sub-branches. Because of its simplicity, the regression tree technique published by (Breiman et al., 1984; Quinlan, 1993) is an automated machine learning model that is frequently used in data mining (Wu and Kumar, 2009). Calculate the standard deviation (Standard Deviation) and assess the dispersion of a data set using the regression tree technique. A big standard deviation indicates that the data has a high degree of dispersion and variability.

The symbol for Standard Deviation is σ (Sigma)

And then compute the standard deviation by taking the square root of the variance.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

The standard deviation calculation procedure consists of four steps:

1) *Standard.*
2) *Average* and squared results for each number
3) *Next,* compute the mean of the squared differences.
4) *Multiply* all values by the square root.

After calculating the standard deviation, we use the regression tree approach to split the data.

The regression tree algorithm goes through the following steps: [11]

1) *Begin* with a single node that contains all of the points. (Calculate standard deviation using the formula)
2) *Stop* if all node points have the same value for all independent variables. Searching for a variable over all binary divisions of all variables, on the other hand, will lower Sigma as much as feasible.
3) *Repeat* step 1 for each new node.

We proceed to analyse the model when we have obtained the findings.

E. Equation Methods

The RMSE, MSE, and MAE indices are often used to evaluate the accuracy of a regression issue. The standard deviation of the residuals (the prediction error) is the Root Mean Square Error (RMSE). The residual is a measure of how far the data points are from the regression line; the RMSE is a measure of how diffuse these residuals are. In other words, it indicates how dense the data is around the best-fit line. In climate research, forecasting, and regression, the base mean square error is frequently used to validate experimental results. The RMSE assessment technique is utilised in this work according to the formula [33]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - r_i)^2}$$

The independent variable is pi, and the predicted value is ri.

When the system is performing optimally, these measurement values approach zero. The greater this value, the poorer the system's efficiency.

IV. METHODOLOGY

A. Datasets

The data represents the sales results for the three years 2017, 2018, and 2019 from a company that specializes in selling electrical and electronic equipment on a national basis. With almost 10 million goods distributed annually. Includes product groups: Home entertainment equipment; Household products; Kitchen products; Air conditioning; health and beauty support equipment, etc. Predicted data is the result of the number of goods sold each year. Table I shows some information of the data.

TABLE I. SOME INFORMATION ABOUT INDUSTRY CODES IN 2018

No	Product code (explanation)		Quantity
1	BEAUTY	Beauty equipment	36
2	C-BATT	The battery	60
3	COLDCHAIN	Freezer	46
…	……	……	…
26	TELEPHONE	Phone	34
27	VC	Vacuum cleaner	17
28	WM	Washing machine	56

B. Algorithm Diagram

The research model consists of two stages in Fig. 2: Stage 1: using the "clustering" algorithm to cluster data according to 4 main characteristics: product type, time, sales quantity and product consumption location. Phase 2: use the prediction method using the "Regression tree" algorithm to build a system to predict product development trends/sales.

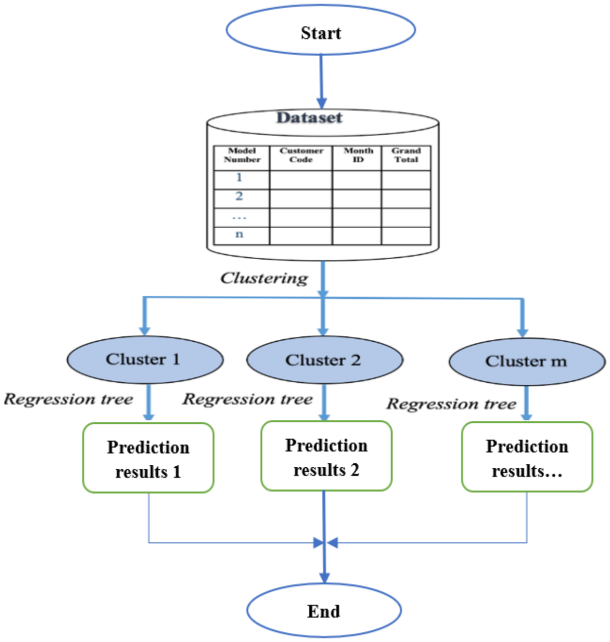


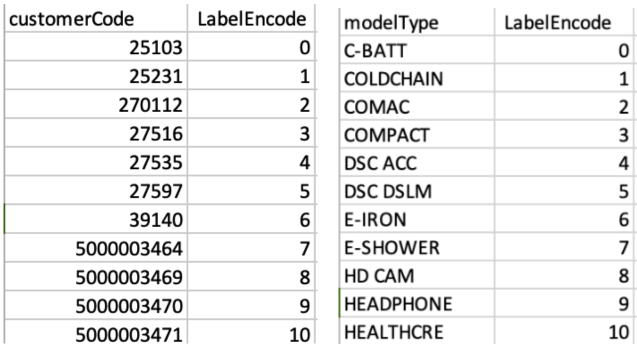
Fig. 2. Algorithm diagram.

C. Data Preprocessing

In practise, data is frequently heterogeneous in terms of data type, with data consisting of several properties (columns) with varying units and magnitudes. This has an impact on the efficiency of many algorithms, as well as their correctness. As a result, before used, the data must be normalised. When the data domains of mixed attributes differ substantially, normalising LabelEncoder data from the sklearn package is a technique widely used as part of preprocessing. The purpose of normalisation is to convert the values of discrete text-valued columns in the data set to a common scale while preserving the range of values. Because of the discrete nature of the data, it is necessary to normalise the values of attributes such as commodity code (Model Type), customer code (Customer Code), month code (Month ID), and total number of items sold (Grand Total) on a common scale in order to generate predictions. Fig. 3 shows the Numeric Data type after standardization, and Fig. 4 is represented as a quantity distribution chart by product type.

D. Clustering (State 1)

With a data set of goods that includes 30 product groups and four different attributes such as commodity code (Model Type), customer code (Customer Code), month code (Month ID), and total number of products sold (Grand Total), the algorithm creates the centre value or calculates the distance in turn before moving on to the next group.



customerCode	LabelEncode
25103	0
25231	1
270112	2
27516	3
27535	4
27597	5
39140	6
5000003464	7
5000003469	8
5000003470	9
5000003471	10

modelType	LabelEncode
C-BATT	0
COLDCHAIN	1
COMAC	2
COMPACT	3
DSC ACC	4
DSC DSLM	5
E-IRON	6
E-SHOWER	7
HD CAM	8
HEADPHONE	9
HEALTHCRE	10

Fig. 3. Data after normalization in numeric form.

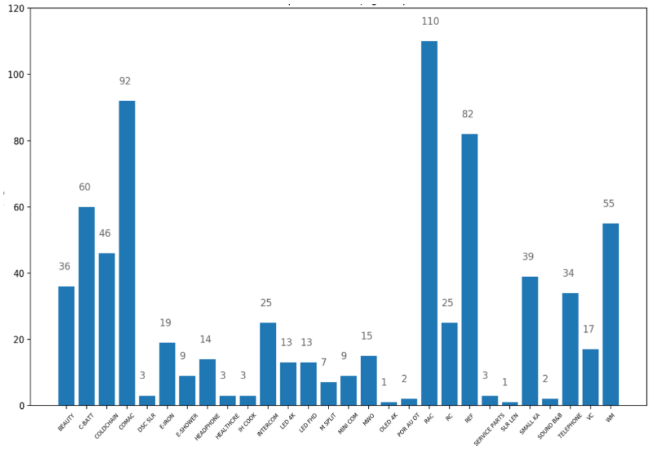


Fig. 4. Quantity distribution chart by product type in 2018.

The clustering algorithm's initial step is to employ k-means [10, 22, and 15]. We set the initial centre values for the groupings. If the data is clustered into k clusters, the kmeans_init_centers function chooses a starting point at random from the data set (Model Number, Customer Code, Month ID, Grand Total).

Specifically, dA(i,j) represents the distance based on the Month ID and Total Grand values, whereas dC represents the distance based on the Model Number and Customer Code characteristics. d(i,j) is determined using the Euclidean distance and two random numbers i and j.

$$\mathrm{dA(i,j)} = \sqrt{(|x_i - x_j|)^2}$$

Because the K-means method only works on numeric data and the remaining characteristics have mixed data types, the K-Prototypes algorithm should be used as shown below.

$$dC(i,j) = \frac{p-m}{p}$$

where, m is the total number of variables (attributes) and p is the number of matches

After determining the two distances dA and dC, compute the common distance for two values of i and j as follows.

$$d(i,j) = m * dA(i,j) + (1-m) * dC(i,j)$$

This formula will return values that are comparable. The distribution of items in the group is extremely skewed in certain circumstances where there are no elements in the group. This problem is solved by repeating the clustering procedure until no empty groups remain.

E. Predict (State 2)

Building a system to anticipate the development trend / product sales using the "Regression tree" algorithm's prediction approach.

After clustering, the findings are discovered to be groupings of cluster data aspects of products, with each cluster having unique qualities. The user will choose which cluster he or she belongs to before applying the Regression Tree issue to each cluster to compute the likelihood of correctly forecasting the quantity of goods sold [10].

Algorithm: Regression tree

Input: - Dataset: data set S each cluster (cluster 1, cluster 2, cluster 3)

Output: - Result tree of each cluster

```
- S = {$s_1$, $s_2$, …$s_k$} is a set Sigma.
- Initialize the set of all points: Sk: f(x) = {x1, x2, …xn}. xi are the points in the set S with N elements.
    Initialize the set f(yj): y = f(xi) is the points xi with the same value.
    If j = 1:
        End
        Calculate the standard deviation of each group f(yj) (yj is the jth element in the set f(y))
        $\sigma$ (f($y_j$)) = Sqrt (1/N* Σ ($x_i$ – $\mu$)$^2$)
        Calculate the standard deviation of the set f(y) and f(x).
    $\sigma$ (f(y)) = Sqrt (1/N *Σ($y_i$ - $\mu_y$)$^2$) …
    $\sigma$ (f(x)) = $\sigma$ j→n (j/N * $\sigma$ (f($y_j$))
        Partition by RSS:
        RSS($s_k$) = $\sigma$ (f(y)) - $\sigma$ (f(x))
Go back to step 2 with the next k.
Evaluating the new model with the largest RSS(sk) will have the most sigma reduction.
- End
```

F. Evaluation

To develop predictive models, we use data from 2017, 2018, and 2019.

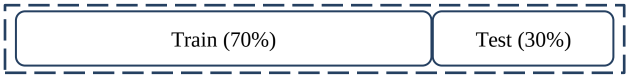


Fig. 5. Evaluation methodology illustration.

And, to assess the accuracy of the prediction results, it is advised to research using the MSE and RMSE formulae to determine the overall mean error with a big quantity of data and the prediction results having a substantial departure from reality. To test the accuracy, the user compares the measured value to the real business value of the firm in 2019. The prediction model's efficacy may then be validated, is shown through the evaluation methodology (see Fig. 5) and flowchart (see Fig. 6).

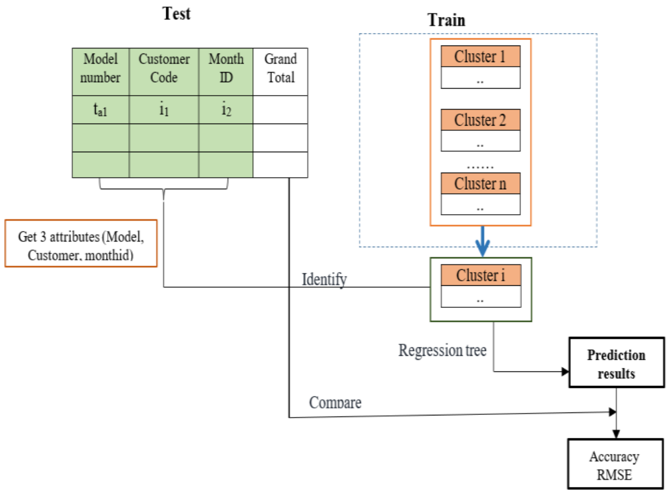


Fig. 6. Flowchart of the test protocol execution process.

V. Results

A. Results of Clustering on the Data Set

Before employing clustering methods, the number of groups k appropriate for clustering must be determined. Choosing k becomes simple for data sets with a modest number of components. Choosing k in a data collection with a high number of components is similarly tough. Combine using the Elbow approach, and then proceed to pick the best k possible using the Elbow chart, which shows the relationship between the SSE value and the number of k groups, with k ranging from 1 to 9 (see Fig. 7). The graph shows a considerable bend at the value k = 3, indicating that the value k = 3 is the optimal number of clusters.

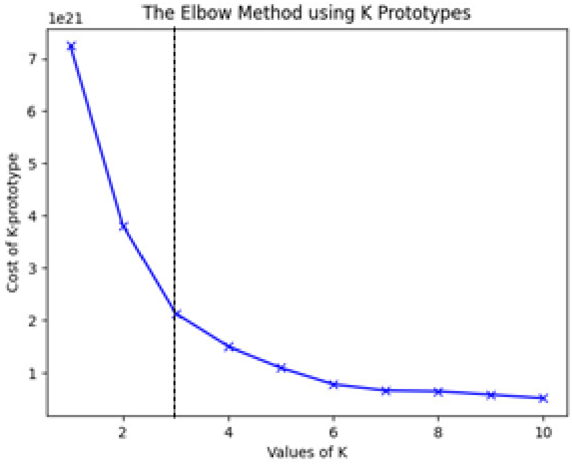


Fig. 7. Elbow method results on specific data set.

After determining the value of K=3, the study conducted clustering for the data set.

The number of items in each cluster can fluctuate as the centroids of the clusters change with the matching k centroids from Table II. With the data set comprising two types of characteristics, numeric attributes and discrete attributes, the result of the centre value of each cluster has two types matching to the items in each cluster with k = 3. Fig. 8 depicts the information of the components in the appropriate cluster when k = 3.

TABLE II. INFORMATION ON CLUSTERS MATCHING TO THE CENTRE K

Cluster number k=3	Cluster	Number of elements
[[2.01743391e+05, '5000003626.0','RAC', 3.32668862e+09], [2.01777722e+05, '5000003903.0','REF', 1.83788563e+07], [2.01765167e+05, '5000003626.0','RAC', 4.00993996e+08]]	1	771422
	2	11874
	3	312

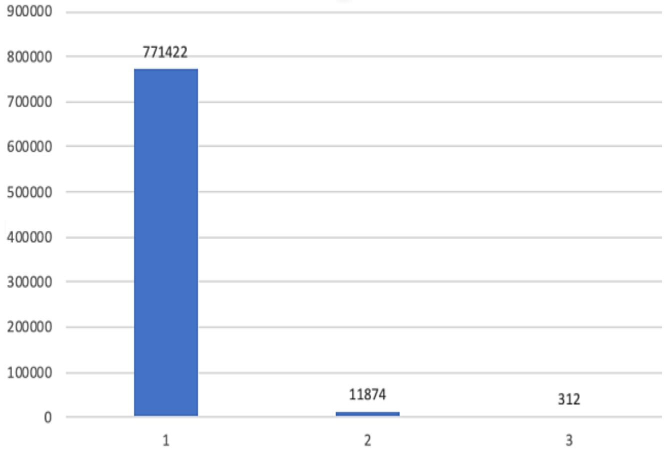


Fig. 8. The graph shows information of each corresponding cluster k=3.

B. Prediction Results on the Whole Data Set

For each cluster, the prediction results are presented on the regression tree.

Due to the big data set, when the tree partition is extremely large, the study reveals that the tree depth is five levels, and the findings mostly demonstrate the partitioning attribute at the root node for each cluster displayed via tree diagrams (Fig. 9, 10, 11).

Cluster 1:

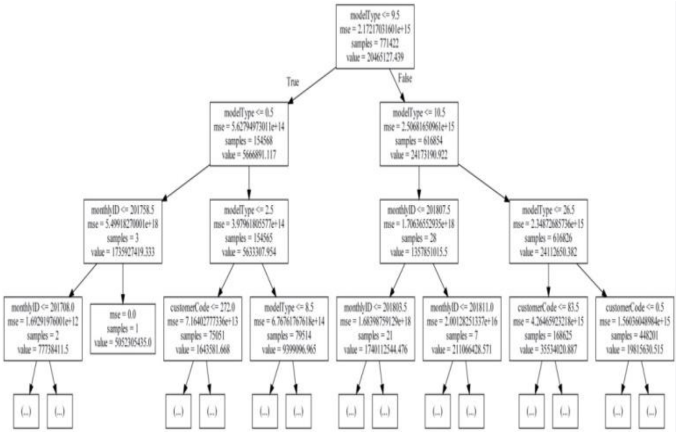


Fig. 9. The tree diagram displays information within the cluster.

Cluster 2:

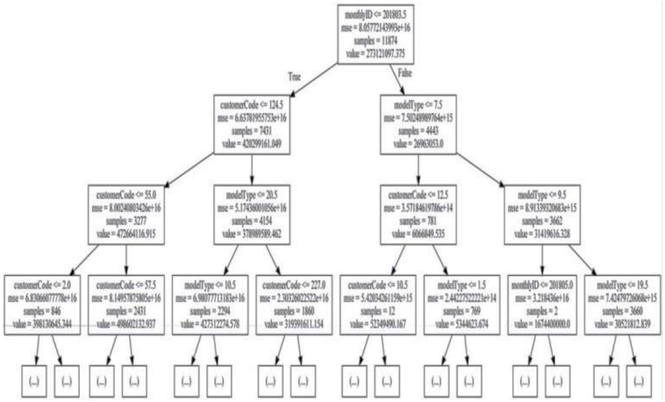


Fig. 10. The tree diagram displays information within the cluster.

Cluster 3:

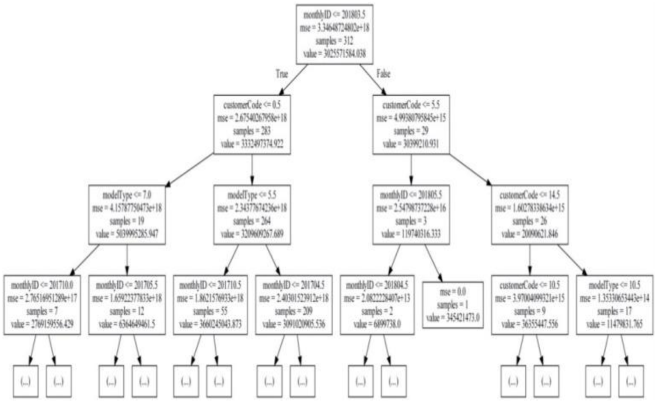


Fig. 11. The tree diagram displays information within the cluster.

The results of splitting the tree by cluster demonstrate that the data on each cluster is dispersed by distinct product groups.

Forecast results are exhibited quarterly throughout the year in the form of a column chart (Fig. 12, 13, 14, 15); the results reveal that the expected value is sometimes high and sometimes low owing to the real scenario caused by weather in each quarter and consumer wants changes, the value will change correspondingly.

Precious 1:

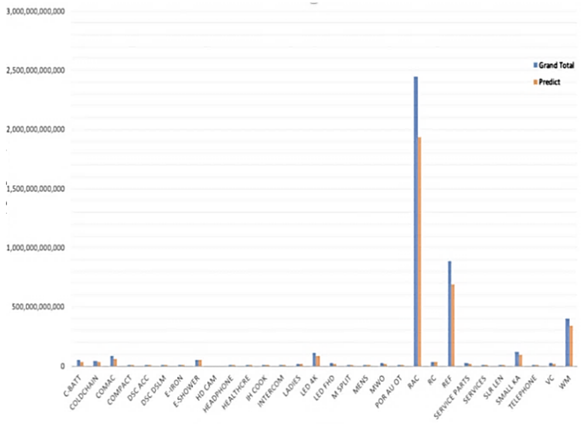


Fig. 12. Predicted and actual results on a specific Q1 data set.

<u>Precious 2:</u>

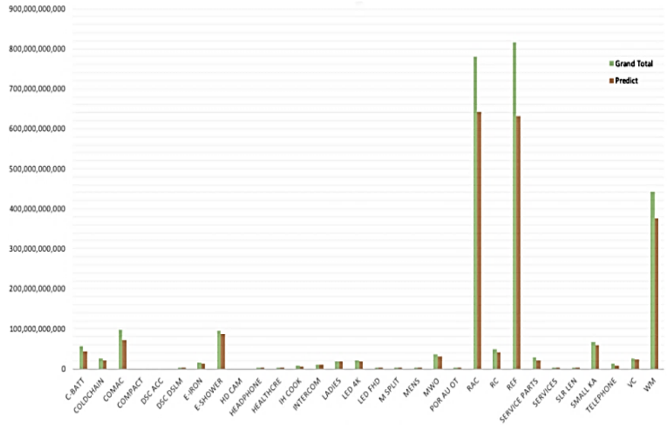


Fig. 13. Predicted and actual results on a specific Q2 data set.

<u>Precious 3:</u>

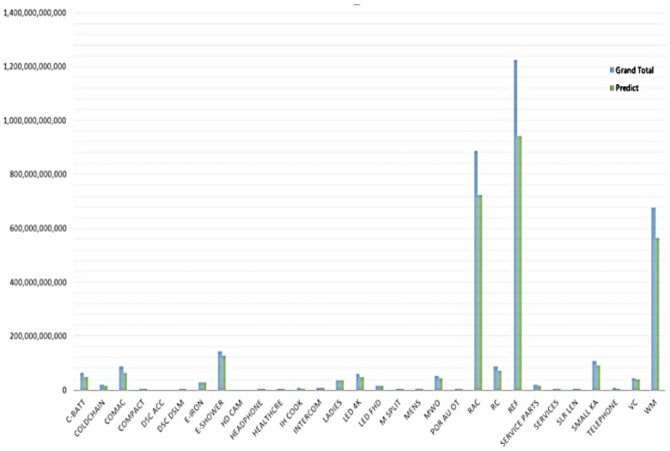


Fig. 14. Predicted and actual results on a specific Q3 data set.

<u>Precious 4:</u>

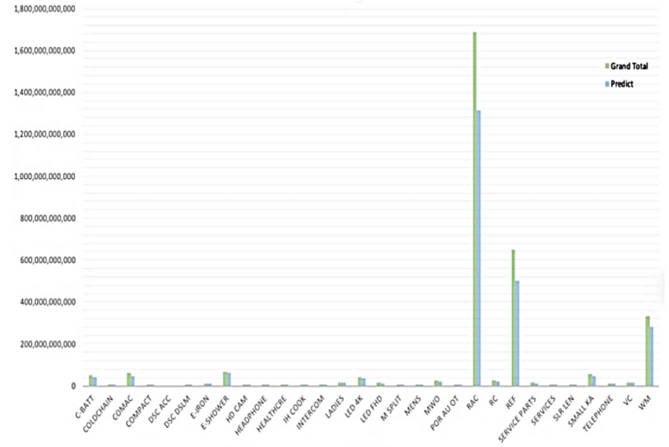


Fig. 15. Predicted and actual results on a specific Q4 data set.

Due to the enormous data collection, the research values are given on each quarter of the year to clearly highlight the difference of each separate product group.

The results of computing the standard deviation error of two cases when forecasting on a test data set based on real data.

TABLE III. DEVIATION RESULTS FOR EVALUATION INDICATORS

	RMSE	**Execution time**
When clustering is note performed	1.163.173,458	45 minute
After performing clustering	43.065,657	40 minute

According to the results of Table III, the deviation and implementation time are lower when clustering before predicting than when not clustering. As a result, the experimental model fits the requirements.

VI. CONLUSION

When the amount of information is really large, the sales forecasting approach is a very useful method for users. It titled "Application of data mining to forecast business trends for electronic products". Proposed clustering model paired with prediction algorithm to be used to the process in order to provide reliable prediction results. The data collection covers product information for three years. Divide the dataset into two parts: the training set contains data from 2017-2018, and the test set contains data from 2019. Using the training dataset with data characteristics of two types of mixed and numeric attributes, research will be conducted based on four attributes: product group (ModelType), customer code (Customer number), month code (Month id), and total quantity of goods sold based on the above three factors. With 3 clusters (k = 3), the number of elements in each cluster is 771422, 11874, and 312, respectively; Combined with the regression tree algorithm "Regression tree" to partition each cluster, using the evaluation protocol MSE, RMSE to evaluate the model's accuracy with over 80% results compared to the actual value, in order to build a system to predict product development trends/sales in the coming years. The experimental findings clearly indicate that using the sales prediction approach in a machine learning programming language yields results with an accuracy of more than 80%. The experimental time, whether rapid or slow, is determined by the original data collection. With the results of the experiment, it is feasible to apply a thorough test data set for each product group of particular categories on additional prediction models such as linear regression, Bagging, and so on in the future. Futures might be based on the outcomes of forecasts put into compact application.

REFERENCES

[1] Mehmet Yasin OZSAGLAM, “DATA MINING TECHNIQUES FOR SALES FORECASTINGS”, (September, 2015), PP. 6-9.

[2] Alfiah, Fifit et al. “Data Mining Systems to Determine Sales Trends and Quantity Forecast Using Association Rule and CRISP-DM Method.” (2018).

[3] Aditya Joshi et al, “Use of Data Mining Techniques to Improve the Effectiveness of Sales and Marketing”. IJCSMC, Vol.4 Issue.4, April-2015, pg. 81-87.

[4] Bhavin Parikh et al, “Applying Data Mining to Demand Forecasting and Product Allocations”. (2003).

[5] Ismail, Mustapha et al. “Data Mining in Electronic Commerce: Benefits and Challenges.” (2015).

[6] Ismail, Mustapha et al. “Data Mining in Electronic Commerce: Benefits and Challenges.” (2015).

[7] Do Thanh Nghi, Le Thanh Van Textbook of Knowledge Systems and Data Mining. Can Tho university, 2012.

[8] Do Thanh Nghi và Pham Nguyen Khang, 2012. Textbook of Principles of Machine Learning. Can Tho University, 137 pages.

[9] Do Thanh Nghi, Python Programming Language. Can Tho University, 2016.

[10] Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984). Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8.

[11] Cosma Shalizi. Statistics 36-350: Data Mining, Fall 2006.

[12] Vu Huu Tiep, Text book Machine Learning co ban, June 15, 2019.

[13] Magidson, J & Vermunt, JK 2002, 'Latent class models for clustering: a comparison with K-means', Canadian Journal of Marketing Research, vol. 20, no. 1, pp. 36-43.

[14] Hendrickson, J. L.(2014). Methods for Clustering Mixed Data. (Doctoral dissertation). Retrieved from.

[15] L. Breiman, Bagging predictors, Mach. Learn, vol. 24, no. 2, pp. 123-140, 1996.

[16] Bühlmann, Peter. (2012). Bagging, Boosting and Ensemble Methods. Handbook of Computational Statistics. 10.1007/978-3-642-21551-3_33.

[17] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[18] NCSS Statistical Software, Chapter 446, [446-1:446-9].

[19] G. W. Milligan and M. C. Cooper. “An examination of procedures for determiningthe number of clusters in a data set”. Psychometrica, 50:159–179, 1985.

[20] T. Calinski and J. Harabasz, “A dendrite method for cluster analysis. Communications in Statistics”, 3:1–27, 1974.

[21] D T Pham, S S Dimov, and C D Nguyen, “Selection of K in K-means clustering”, 2004.

[22] Andrew Ng, “Clustering with the K-Means Algorithm, Machine Learning”, 2012.

[23] Robert L. Thorndike (December 1953). “Who Belong in the Family?”. Psychometrika 18 (4): 267–276.

[24] Karolis Urbonas, “Practical implementation of k-means clustering”.

[25] Peter J.Rousseeuw, “Silhouettes: a graphical aid to the interpretation and alidation of cluster analysis”, 1986.

[26] Prakash Nadkarni ,“Chapter 10 - Core Technologies: Data Mining and “Big Data””, 2016.

[27] Bednarz T. F., (2011): Sales Forecasting: Pinpoint Sales Management Skill DevelopmentTraining Series, Majorium Business Press, p.46.

[28] Kaufman L. and Rousseeuw P.J: Finding groups in Data. An introduction to cluster analysis. Wiley Interscience, 2005.

[29] Gatu, C., Yanev P.I., Kontoghiorghes, E.J., 2007. A graph approach to generate all possible regression submodels. Comput. Statist. Data Anal., 52, 799–815.

[30] Hofmann, M., Gatu, C., Kontoghiorghes, E.J., 2007. Efficient algorithms for computing the best subset regression models for large-scale problems. Comput. Statist. Data Anal., 52, 16– 29.

[31] Shih Y.S., Tsai H.W.,2004. Variable selection bias in regression trees with constant fits, Comput. Statist. Data Anal., 45, 595–607.

[32] A Maesya and T Hendiyanti, “Forecasting Student Graduation With Classification And Regression Tree (CART) Algorithm”, 2018.

[33] Barnston, A. G., 1992: Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score. Wea. Forecasting, 7, 699–709.

[34] Clustering Algorithms. By John A. Hartigan. New york: John Wiley and Sons, 1975.

[35] D T Pham, S S Dimov, and C D Nguyen, “Selection of K in K-means clustering”, 2004.

[36] Z. Huang, “Clustering large data sets with mixed numeric and categorical values,” in Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining, Singapore, 1997.

[37] Murty, M.. (2013). Cluster Analysis on Different Data Sets Using K-Modes and K-Prototype Algorithms. 249. 10.1007/978-3-319-03095-1_15.

[38] Özlem Akay & Güzin Yüksel (2018) Clustering the mixed panel dataset using Gower's distance and k-prototypes algorithms, Communications in Statistics - Simulation and Computation, 47:10, 3031-3041, DOI: 10.1080/03610918.2017.1367806.

[39] Dake, Delali & Gyimah, Esther & Buabeng-Andoh, Charles. (2023). University Students Behaviour Modelling Using the K-Prototype Clustering Algorithm. Mathematical Problems in Engineering. 2023. 10.1155/2023/5507814.

[40] Bitzenis, Aristidis & Koutsoupias, Nikos & Boutsiouki, Sofia. (2023). Business Research and Data Mining: a Bibliometric Analysis. 10.1109/ICECCME57830.2023.10252699.

[41] Alice, Dr & Andrabi, Syed & Jha, Shambhavi. (2023). Sales Forecasting Based on Ensemble Learning. 10.36227/techrxiv.24049452.v1.

Identification of Microaneurysms and Exudates for Early Detection of Diabetic Retinopathy

G. Indira Devi[1], D. Madhavi[2]
Electronics and Communication Engineering, Anil Neerukonda Institute of Science and Technology, Visakhapatnam, India[1]
Department of ECE, GIT, GITAM University, Visakhapatnam, India[2]

***Abstract*—Diabetic retinopathy (DR) is a condition that may be a complication of diabetes, and it can damage both the retina and other small blood vessels throughout the body. Microaneurysms (MA's) and Hard exudates (HE's) are two symptoms that occur in the early stage of DR. Accurate and reliable detection of MA's and HE's in color fundus images has great importance for DR screening. Here, a machine learning algorithm has been presented in this paper that detects MA's and HE's in fundus images of the retina. In this research a dynamic thresholding and fuzzy c mean clustering with characteristic feature extraction and different classification techniques are used for detection of MA's and HE's. The performance of system is evaluated by computing the parameters like sensitivity, specificity, accuracy, and precision. The results are compared between different types of classifiers. The Logistic Regression classifier (LRC) performance is good when compared with other classifiers with an accuracy of 94.6% in detection of MA's and 96.2% in detection of HE's.**

Keywords—Diabetic retinopathy; microaneurysms; hard exudates; SVM; LRC

I. Introduction

For the last fifty years, diabetic retinopathy (DR) has been considered the most prevalent cause of blindness. According to epidemiological research conducted in developed nations, DR is one of the four primary causes of vision impairments in the general population. Diabetes damages the blood vessels of the human retina, which is one of the primary causes of visual impairment in DR [1]. In contrast to the nearly non-occurrence of DR in the first five years following a type I diabetes diagnosis, latest is type II which is found in a ratio of 1:5 i.e., one diabetic among five tests has DR at the time of diagnosis. Nonetheless, nearly all type I diabetic patients and two-thirds of type II diabetic individuals have DR symptoms after 15 years.

The most typical signs of DR are abrupt loss of vision, floaters and flashes, and impaired vision. Since DR is a progressive illness, whose severity is dictated by the quantity and variety of lesions visible in the fundus picture, early identification and treatment are essential. The macula, optic disc, and blood vessels that make up a healthy retina are its key constituents; any alterations to these elements are indicative of an eye illness [2]. Non-proliferative DR (NPDR) and proliferative DR (PDR) are the two main phases of DR. NPDR, often referred to as background DR, is a condition in which diabetes destroys the blood vessels in the retina, allowing fluid and blood to seep onto the surface of the retina [3].

Due to the process of leaking the retina gets moist, swollen and finally loses functioning. A variety of retinopathy symptoms, including microaneurysms (MAs), hemorrhages (H), hard-exudates (HE), soft-exudates, or cotton wool spots (CWS), may be present in NPDR. Three phases of non-primary depression (NPDR) are distinguished depending on the severity and number of lesions. These stages included in DR is mild, moderate, and severe which are caused by localized dilatations of thin blood arteries. MAs are the initial indication of NPDR. MAs are tiny, nearly circular, and have a crimson colour H, often known as dot or blot H, is the next indication of DR. Thin vessel or MA walls that are sufficiently weakened may burst and result in H. Larger red lesions are seen in blot haemorrhages, whereas brilliant little red spots are seen in dot haemorrhages. Dot hemorrhages and MAs are occasionally grouped together as a single red lesion type called HMAs. For convenience of viewing, HMA-containing areas in the retinal picture are magnified in Fig. 1.

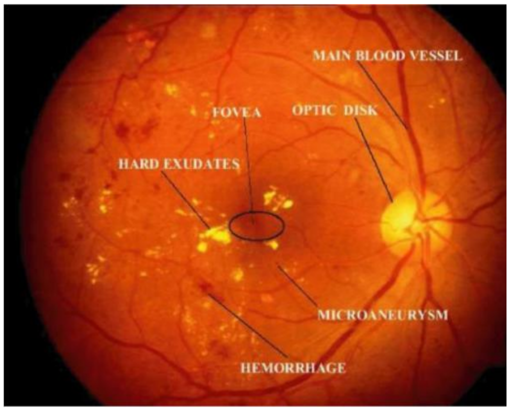


Fig. 1. Retinal fundus imgae shwoing MA's and HE's.

When fundus imaging is used to diagnose DR, microaneurysms (MAs) are the initial pathological signs that are identified. They manifest as minute blood bulges that resemble little red spots on the retina. Retinal blood vessels leak due to microaneurysms. Lipids and fluids leak from blood vessels as the condition worsens, forming hard exudates that are yellowish in colour and come in a different type of shapes and sizes. The macula and fovea are the regions in charge of central vision; if exudates accumulate there, the patient may lose or have degraded central vision. In automated DR detection, the Hard Exudates' detection is crucial. To find MAs and HEs, a machine learning method has been created. In this work both MAs and HEs are detected at a time for the given input image. The need of identifying each of them individually is avoided. So that the treatment can be provided effectively utilizing single process.

The proposal of work is organized in different sections. In Section II a brief overview of existing methods in detection of MA’s and HE’s is been discussed. The proposed method of extracting features and classification model is discussed in Section III. The experimental findings are evaluated in Section IV and results for identification are given in Section V. Finally, the overall summary is given in Section VI.

II. Related Work

The diagnosis of DR has made extensive use of computer-aided automated analysis of colour fundus pictures, which is very effective [4]. This might lessen the burden on ophthalmologists and increase the effectiveness of DR screening. Due to the significance of MA for DR diagnosis and the advancement of computer-aided diagnosis (CAD), an increasing number of research on automatic MA detection have been conducted recently. The author in study [5] used the Support Vector Machine (SVM) and Naive Bayesian (NB) classifier to distinguish between red and brilliant lesions; however, they were only able to use one database with 100 images for training and testing. The author in study [6] developed a Random Forest (RF) based approach for the diagnosis of both MA and HM using dynamic shape data, without the need for any prior segmentation of lesions. However, the blood vessel-related lesions were overlooked which resulting in false-negatives (FN) [7].

In research [8], the author suggested filters with various grid sizes combined with MKL, SVM, and filters to cope with false-positives (FP) brought on by tiny blood vessel (BV) segments to detect MA and HM. It was discovered that employing MKL improved performance when compared to using a single grid size, although choosing a larger grid size came with a greater computing cost. To get over the issue of class imbalance, the author in study [9] offered an unsupervised classification approach for MA identification based on sparse principal component analysis (PCA). However, several FPs occur when extracting features. The author in [10] used peak detection and region expansion to obtain MA candidates, and then used K-nearest neighbours (KNN) to identify MA. This led to an e-ophtha MA database FROC score of 0.273, which is rather low. Singular spectrum analysis (SSA) and a KNN classifier were used by the author in [11] to create a method for MA recognition; however, because there was no subtle or low contrast or blurry-outlined MA, this approach produced multiple false positives (FPs). A few MAs were also overlooked during the selection of candidates.

The author reported a multi-stage automated approach in [12] for identifying longitudinal retinal changes produced by small red colour lesions known as dot HM and MA. The author of [13] employed SVM for classification and local binary pattern (LBP) to extract textural features to detect MA. The author in [14] was able to detect MA by utilizing discriminative dictionary learning (DDL) and multi-feature fusion dictionary learning (MFFDL), respectively. The former mainly relied on the original grayscale feature dictionary, and employing a grayscale feature with single component will impact the performance since retinal pictures vary greatly in terms of colour, luminance, and contrast.

A key factor in the diagnosis of DR is biomedical engineering. Image processing was used in several studies on diabetic retinopathy to analyze retinal images [15]. In diabetic retinopathy, the GMM classifier was used to automatically detect red lesions [16]. To localizing exudate in the colour fundus picture, morphological operators were employed [17]. Furthermore, a few researchers have looked at the retinal image's hard exudate detection for use in the diagnosis of DR. To detect hard exudate, the retinal picture was subjected to a linear brightness adjustment [18]. In the CIE lab colour model, the hard exudates were identified using k-mean on the colour retinal picture [19]. In DR, the exudates were automatically detected using the wavelet transform [20].

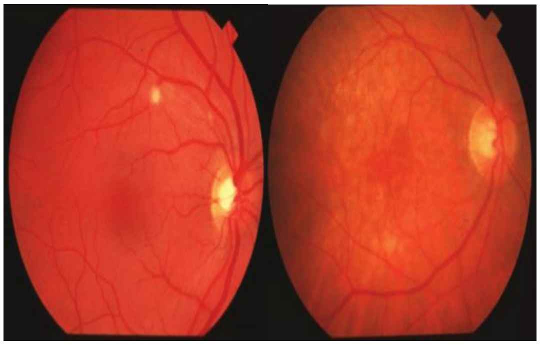
Subsequent research concentrated on deep learning techniques, primarily for image categorization. Convolutional neural networks (CNNs) have the ability to identify picture patches [21] and particular pixels as either exudates or non-exudates [22]. Deep neural networks can be developed from scratch [23] or based on a variety of designs that have been pre-trained on diverse datasets [24] (transfer learning approach). We made the decision to integrate an SVM classifier with transfer learning techniques in order to increase the overall process' efficacy [25].

When it came to DR, there were several methods and procedures for identifying MAs and HEs. The work progressed using morphological processes to identify MA, along with distinctive feature extraction, hessian analysis, and classification algorithms. Dynamic thresholding and fuzzy c mean clustering with distinctive feature extraction are used to identify HE, and classification algorithms are then used.

III. Material and Methods

A. Data Utilized

The proposed methodology's execution is carried out utilizing the STARE dataset. There are 400 pictures in this collection, including HEs, MAs, and other diseased features including red lesions of different sizes and forms. The samples images which are available in STATE dataset are shown in Fig. 2.



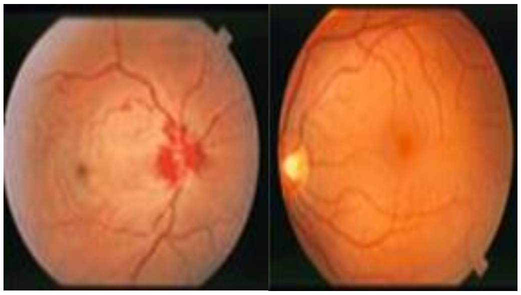


Fig. 2. Images of stare dataset.

The images of the created database were taken with a Topcon camera, which has a 605 x 700 resolution. Each retinal picture is subjected to an image-based DR detection assessment; however, reference markers for evaluating the approach based on the number of exudates-based assessment are absent. Therefore, an eye expert marked the contours of the lesions of 96 retinal pictures with various features to evaluate the exudates-based effectiveness of the suggested technique. The detection process implemented in this work is shown in Fig. 3.

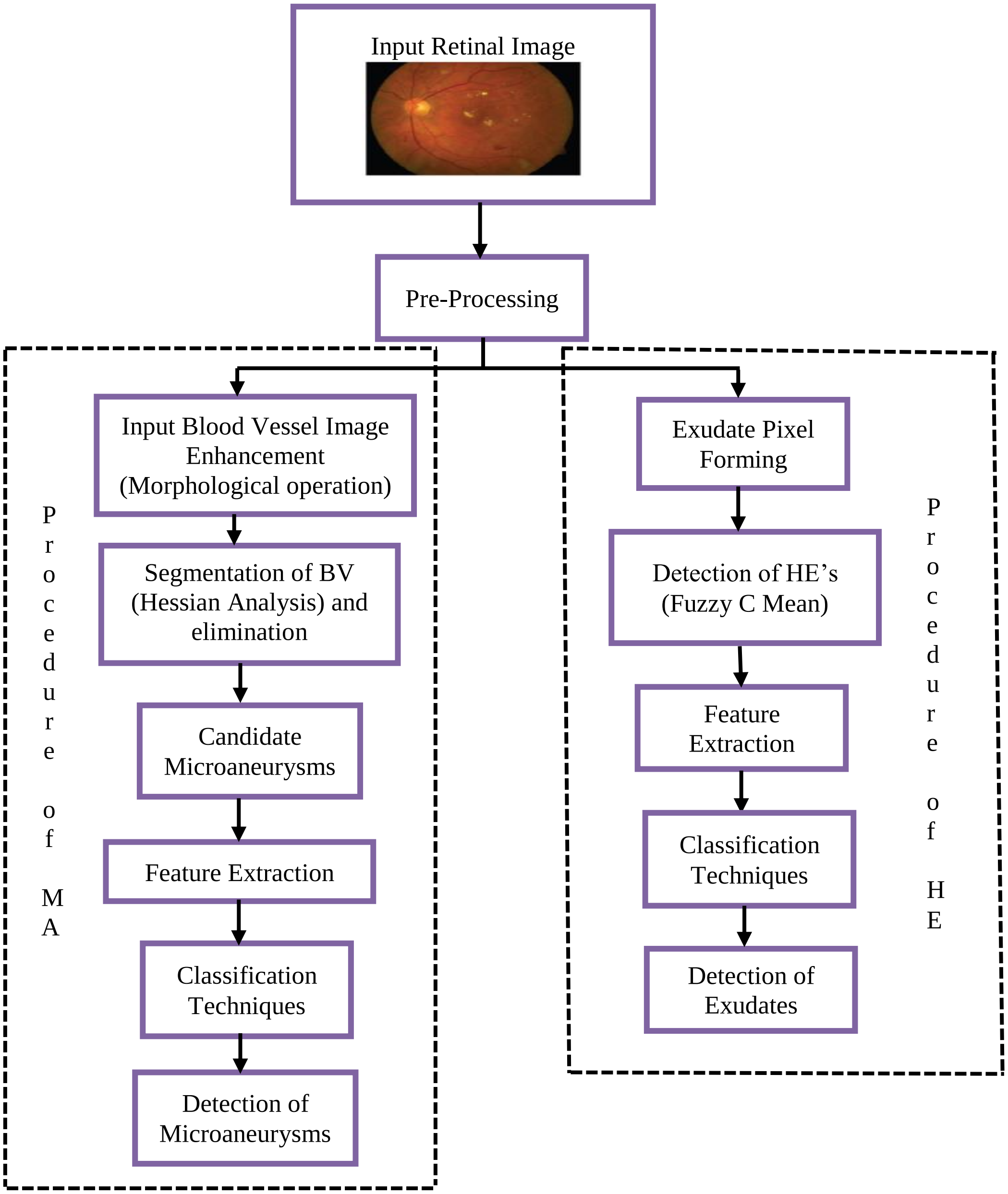

Input Retinal Image
Pre-Processing
Input Blood Vessel Image Enhancement (Morphological operation)
Segmentation of BV (Hessian Analysis) and elimination
Candidate Microaneurysms
Feature Extraction
Classification Techniques
Detection of Microaneurysms
Procedure of MA
Exudate Pixel Forming
Detection of HE's (Fuzzy C Mean)
Feature Extraction
Classification Techniques
Detection of Exudates
Procedure of HE

Fig. 3. Process flow of proposed model.

B. Pre-Processing

The methods in various DR problems are demonstrated using retinal pictures from the STARE dataset. They do, however, differ in brightness and contrast, which are influenced by the surroundings when a fundus camera takes the picture. It causes the complexity of parameter setting and data evaluation, as was previously noted. As a result, the preprocessing portion of the retinal picture is customised to provide appropriate data for additional analysis. The retinal picture is loaded into the system to begin processing. Eq. (1) is then used to adjust the contrast of the retinal picture, producing an image with more clarity than before.

$$C_M(i) = \frac{\max(i)-\min(i)}{\max(i)+\min(i)} \tag{1}$$

The processing is divided into two sections. The first section, the optic disc of retinal image is detected using the algorithm for automatic localization of optic disc in the fundus image [10]. The second one, the retinal image is extracted into three channels (red, green, and blue). The green channel having the complete details of hard exudates is selected for operation. Subsequently, the data is adjusted from 0 to 1 by normalizing function as Eq. (2).

$$N_{(ij)} = \frac{X_{(i,j)}-\min(X)}{\max(X)-\min(X)} \tag{2}$$

The normalized data of the green channel is subtracted by the acquired optic disc.

C. MA Detection

In the process of MA Detection, initially the blood vessels need to be enhanced and later segmented, finally eliminated.

After eliminating the blood vessels, the remaining area of the fundus images without BV clearly shows the MA's.

1) Input blood vessel image enhancement: For the process of enhancing the image of blood vessels morphological filters is been utilized [26]. The contrast between the vessel structure and the backdrop intensity fluctuations is more noticeable. On the other hand, a closer look into vessel intensities may reveal significant alterations that have the potential to negatively impact the extraction process. We have suggested a morphological filter called the modified morph, which has been applied to a normalised green channel picture, to counteract such alterations. The images in the database which contain thickest vessel width is utilized as reference image for performing morphological operations with the range of pixels 1 and 8 for considering the diameter ranges of vessels width. The vessel diameter scale may be modified to account for changes in picture quality.

The morphological operation has been employed to ascertain the disparity among the input image considered and the opened picture. To acquire the inverse picture, first open the image and then close it. The noise sensitivity of the modified morph implementation results in pixel values in an opened image that are always less than or equal to the input values; under these circumstances, the image which are subtracted have low level intensity fluctuations in the data. The operator 'open' of an image 'I' with structuring element S_o is given by,

$$I_o = I \circ S_o \tag{3}$$

The operator 'close' function of an image 'I' with structuring element S_c is given by,

$$I_c = I \cdot S_c \tag{4}$$

Modified morph operation of an image is given by,

$$I_{mod} = I - (I \cdot S_c) \circ S_o \tag{5}$$

The Eq. (5) shows our modified morph operation in which 'I' is the input green channel image while S_c and S_o stand for the elements of structuring for closing (·) and opening (∘) operators, respectively.

2) Segmentation of BV and elimination: Here, the morphological filter is followed in a novel approach by the hessian matrix to improve the quality of the pictures obtained of the thin and broad vessels. For both thin and wide vessel enhancement, we have independently calculated the second derivative of the picture at two distinct scales. The utilisation of a hessian matrix-based technique has facilitated the separation of broad and thin vessels. The focus is on vessels with varying widths, which are determined by analysing the second order derivative at two distinct scales. The hessian matrix's Eigen values and their difference are being utilized to reduce non-vasculature structure and improve contrast. The directed image I_i and its hessian matrix in the updated coordinates $Cx'y'$is found to be,

$$H' = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 I_i}{\partial x'^2} & \frac{\partial^2 I_i}{\partial x' \partial y'} \\ \frac{\partial^2 I_i}{\partial y' \partial x'} & \frac{\partial^2 I_i}{\partial y'^2} \end{bmatrix} \tag{6}$$

Otsu thresholding is applied individually to broad and narrow vessel pictures because applying it to the entire image at once does not yield useful results. So applied a global threshold to the enhanced picture of the broad vessels and then combined it with the enhanced image of the narrow vessels.

Utilising a modified version of Otsu's technique, geometrical objects and undesired noise are suppressed depending on the vessel structure. Otsu's method is often applied locally or globally over the whole picture to determine a threshold for vessel and non-vessel pixel categorization. Otsu threshold is been applied individually to broad and narrow vessel pictures because applying it to the entire image at once does not yield useful results. A global threshold is applied to the enhanced picture of the broad vessels and then combined it with the enhanced image of the narrow vessels. The segmented blood vessel in which one is thin and other is thick vessel is shown in Fig. 4.

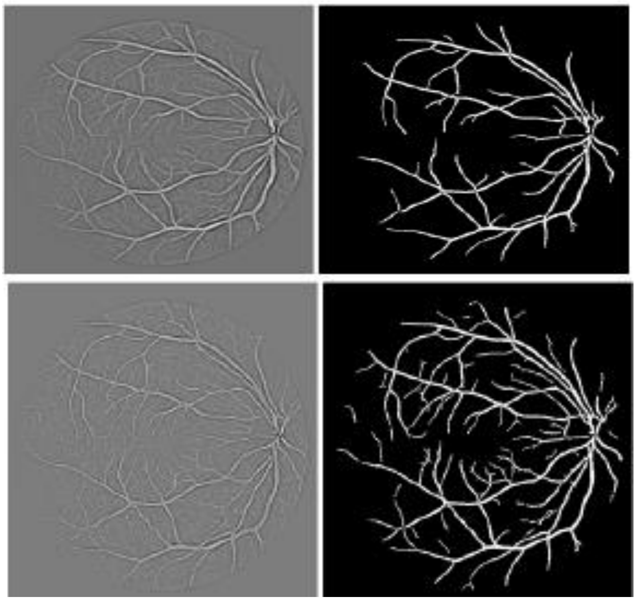


Fig. 4. Thin and thick BVS.

After segmenting the BV's, the blood vessel needs to be eliminated. By eliminating the region of blood vessels, the regions of MA's are easy to extract. The extracted MA region and the morphological filter response on MA region are shown in Fig. 5.

To lessen unwanted noise and geometrical structures according to the design of the vessel, adapted Otsu's technique. Otsu's method is often applied locally or globally over the whole picture to determine a threshold for vessel and non-vessel pixel categorization.

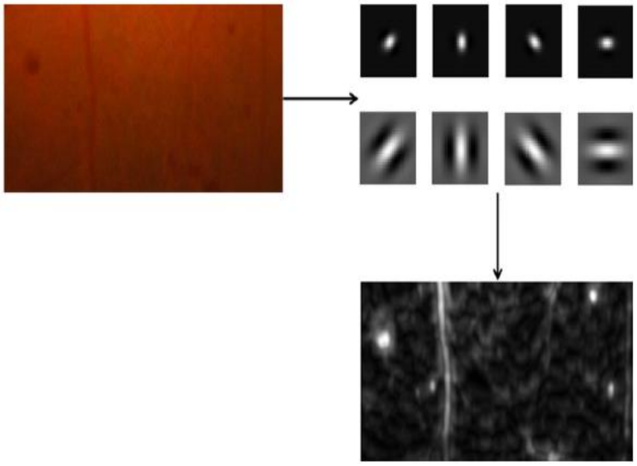


Fig. 5. Slice of Input image with MA region.

3) Feature extraction: All prospective objects (regions) that might be regarded as potential MA regions are provided by the candidate MA region extraction step. The set representation for an image 'n' is if it has 'N' probable candidate areas is given as $n = \{n_1, n_2, n_3, \dots\dots, n_N\}$. A feature vector with all m characteristics, i.e. for a sample lesion, represents each item or potential lesion region n_i, which is regarded as a sample for classification and the feature vector is $n_i = \{fv_1, fv_2, fv_3, \dots\dots, fv_m\}$ where $i = 1,2, \dots\dots, N$.The area, eccentricity, compactness, aspect ratio, mean and standard deviation, mean gradient, mean gradient magnitude, mean HSV (hue, saturation, and value), entropy, homogeneity, and energy value are the properties that we employ in the system which have been presented. Finally, the features are given to different type of classifiers which are discussed in Section III to identify the detection rate of accuracy.

D. HE Detection

1) Exudates pixel forming: Retinal pictures must be cleaned of artefacts in order to improve the quality of retinal fundus images for the precise detection of exudates. These artefacts can be caused by a variety of factors, including focus, blur, and improper illumination of the retinal fundus picture. The quality of retinal pictures is declining due to these reasons. Furthermore, poor picture quality may lead to erroneous retinal disease diagnosis. Therefore, improving photos before determining the presence of exudates is a crucial step.

2) Detection of hard exudates: There were two primary methods. To segment colour retinal pictures, FCM clustering was first utilised to get the local dynamic threshold of each sub-image. This threshold was then coupled with the global threshold matrix. Subsequently, exudates and non-exudates zones were distinguished using classification algorithms.

a) FCM for segmentation of retinal image: The technique of segmenting images using the dynamic threshold in conjunction with the global threshold based on FCM clustering is explained as follows:

i) The retinal picture was split into 'S' sub-images, and each sub-image's pixels were assigned to distinct categories using fuzzy memberships according to the FCM method. The following cost function was minimised by FCM, an iterative optimisation technique:

$$I(A, B) = \sum_{i=1}^{n} \sum_{s=1}^{c} (a_{si})^m \|x_i - C_s\|^2 \quad (7)$$

where, C_s denotes the clustering centre of the kth cluster and a_{si}denotes the membership of pixel x_i in the kth cluster. Since the grayscale value was the sole feature utilised for clustering, the midpoint of the line representing the clustering centre was used as the segmentation threshold, and the sub-image threshold was determined by taking the mean of the two clustering centres;

ii) To determine the global threshold and create the global matrix 'G' with the same dimensions as the original picture, all the pixels in the original retinal image were categorised in the same manner as described previously.

iii) A mean filter with a size of 10 × 10 was applied to the dynamic threshold matrix T_D, which was created by interpolating the thresholds of the individual sub-images into a matrix the same size as the original picture.

iv) The final threshold matrix $T_h$was constructed as,

$$T_h = jG + (1 - j)T_D \quad (8)$$

where, the value of 'j'was set to 0.1.

v) The retinal picture and the threshold matrix T were compared to determine the segmentation outcome. Results of retinal imaging segmentation are influenced by the sub-image's size. The FCM clustering results for various sub-image sizes are displayed in Fig. 6. After considering the accuracy of the local threshold as well as the running duration, 20×35 pixels was determined to be the most appropriate size for the sub-images.

3) Feature extraction: Some notable traits that were frequently employed by eye care professionals to visually identify HE from other forms of lesions were retrieved from each location and used as inputs of classification systems in order to further partition the exudates regions from the exudate's candidates. Area, eccentricity, compactness, aspect ratio, mean and standard deviation, mean gradient, mean gradient magnitude, mean HSV (hue, saturation, and value), entropy, homogeneity, and energy value were among the important characteristics. Lastly, the characteristics are fed into several classifier types—discussed in Section III—to determine the accuracy of the detection rate.

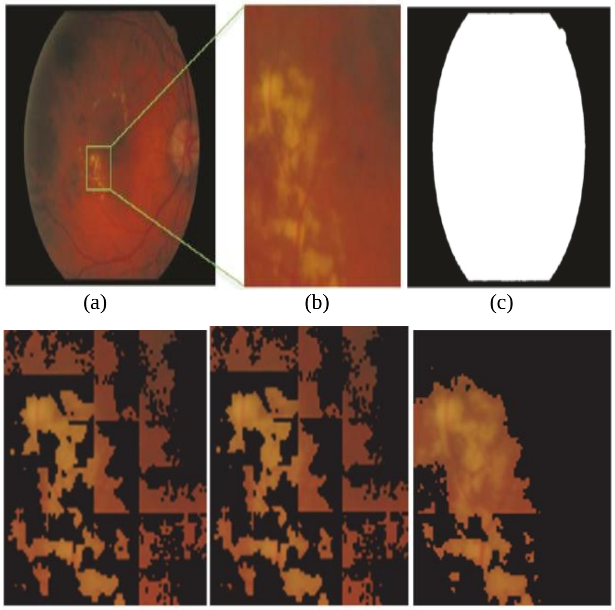


Fig. 6. For different sub-image sizes, the segmentation using FCM clustering. (a) Input image (b) Region of exudates (c) FCM segmented output (d) Segmentation with a size of 10×15. (e) with20×35pixels. (f) With 50×60pixels.

E. Classification Techniques

Following the extraction of the features, five classifiers—DT, RF, SVM, KNN, and LR—are used to begin the classification process. Adaptive threshold algorithms are used in both methods to improve pictures, with 94.6 and 96.2 accuracy levels for MA and HE identification, respectively.

1) Decision Tree (DT): DT is a popular hierarchical machine learning technique for prediction that uses a model of decisions and possible outcomes that resembles a tree. Every branch of DT reflects the test results, every internal node and that is not a leaf node evaluates an attribute, and every leaf node or terminal node specifies the label class.

2) Random Forest (RF): A Random Forest Classifier, which gathers data from MAs and divides it into two likely classes healthy and unhealthy is the third stage. This method operates in a manner akin to that of the decision tree algorithm. This method discovers the optimal answer by compiling all the decision tree predictions. Every tree is dependent on the collection of random features.

3) K-Nearest Neighbour (KNN): KNN is a supervised machine learning technique that predicts the values of new data items by using "feature similarity". In the classification phase, the unlabelled sample is categorised by designating the class label depending on which of the k training samples that are closest to the query point—where k is a predetermined constant—is the most common. In the training phase, this is accomplished by storing the feature vectors and class labels of the training samples.

4) Support Vector Machine (SVM): SVM is a supervised machine learning algorithm that may be applied to problems related to regression and classification. This method constructs a hyperplane (or a group of hyper-planes) in a high- or infinite-dimensional space to determine the best boundary between the potential outputs. Finding a hyperplane that maximises the isolation of data points from possible groupings is the aim in n-dimensional space. SVMs can manage very large feature spaces because they use over-fitting avoidance, which is independent of feature count.

5) Logistic Regression (LR): For classification situations where the goal is to predict the chance that a given instance will belong to a specific class, supervised machine learning techniques termed logistic regression are typically utilised. The word used to describe the classification techniques that use it is logistic regression. The outcome of a categorical dependent variable is predicted using logistic regression. The outcome must thus be a discrete or category value. It gives the probabilistic values that lie between 0 and 1, as opposed to the exact values of 0 and 1. It might be true or false, 0 or 1, yes or no, etc. It is quite like linear regression, except for how they are used.

Regression issues are solved using linear regression, and classification challenges are resolved using logistic regression. We fit a logistic function with a "S" shape that predicts two maximum values, 0 and 1, in lieu of fitting a regression line in LR. The logistic function's curve indicates a number of potential outcomes, such as the presence of cancerous cells and the relationship between a mouse's weight and fatness.

Algorithm for Training:

This section outlines the fundamental procedures for creating a picture (training set) and a collection of targets, as well as the processes we do in the order listed below:

$$\text{For i} = 1, \text{to the number of training images,}$$

$$\text{Obtain the i}^{\text{th}} \text{ image from the database,}$$

Apply to preprocess and extract the features,

Stack the feature results to the training image array,

Assign a target class based on the severity of the dataset

Train classifiers: The following classifiers are picked to carry out the actions, and this sub segment specifies the necessary procedures to train the chosen classifier:

- Train the DT classifier,
- Train the RF classifier,
- Train the KNN classifier,
- Train the SVM classifier, and
- Train the LR classifier.

Algorithm for Testing:

This section outlines the general processing procedures to forecast the outcomes using the provided classifier's feature extraction technique and the subsequent steps we do in that order:

Step1. Obtain the desired test image,

Step2. Apply image pre-processing and extract the features,

Step3. Predict the DT classifier features

Step4. Predict the RF classifier features

Step5. Predict the KNN classifier features,

Step6. Predict the SVM classifier features

Step7. Predict the LR classifier features,

Step8. Obtain the model of prediction results.

IV. Experimental Findings

To assess the effectiveness of the suggested detection framework, retinal pictures from the publicly accessible STARE dataset were analysed in this study. Expertly segmented retinal pictures constitute the dataset, which is often regarded as the gold standard for comparison. The 40 and 20 retinal pictures in the STARE datasets, respectively, are divided into training and test sets. The suggested methodology has been used to evaluate performance using 20 test photos from the STARE dataset. Our suggested framework's trials were all carried out using the aid of the software programme MATLAB 2021a.

The parameters evaluated to show the performance of proposed detection model are sensitivity, specificity, precision, and accuracy. The finding of accuracy is based on true positive, true negative, false positive and false negative values. Here the values signify how effective the MA’s and HE’s are detected. The classification needs to be correct to achieve correct values. The entire performance of segmentation is shown by accuracy. Specificity measures the identification of pixels with negative values, whereas sensitivity represents the ability of detecting pixels with positive values. The following formulae provide the findings, which are displayed in Tables I and II.

$$Sensitivity = \frac{TP}{TP+FN} \tag{9}$$

$$Specificity = \frac{TN}{TN+FN} \tag{10}$$

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

$$Accuracy = \frac{TP\_TN}{TP+TN+FP+FN} \tag{12}$$

V. Results for MA’s Identification

TABLE I. RESULTS FOR MA’S IDENTIFICATION

Classification Technique	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
Decision Tree	83.4	85.4	84.7	87.5
Random Forest	87.6	88.3	87.2	89.6
KNN	93.7	94.2	94.8	95.2
SVM	94.2	95.7	94.9	96.4
Logistic Regression	94.6	96.5	96.2	97.6

TABLE II. RESULTS FOR HE’S IDENTIFICATION

Technique	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
Math Morph [26]	0.92	0.89	0.91	-
Deep CNN [27]	0.69	0.64	0.88	-
2 step CNN [28]	-	0.77	-	-
MSRNet [29]	-	0.71	-	-
Decision Tree	84.6	86.5	85.2	88.2
Random Forest	87.2	89.1	88.6	90.3
KNN	93.1	95.3	95.1	96.7
SVM	95.2	96.4	95.9	97.8
Logistic Regression	96.2	97.6	97.2	98.4

TABLE III. OVERALL ACCURACY OBTAINED USING PROPOSED MODEL

Classification Technique	MA’s and HE’s Accuracy (%)
Decision Tree	84.0
Random Forest	87.4
KNN	93.8
SVM	94.7
Logistic Regression	95.3

The discussion of Table III is all about the overall rate of accuracy obtained by various machine learning techniques in detection of MAs and HEs for the given STARE dataset. The LR classification achieved higher rate of accuracy with 95.3% when compared to techniques like DT, RF, KNN, and SVM. The classifiers performance may be more accurately assessed using the receiver operating characteristic (ROC) curve and is shown in Fig. 7.

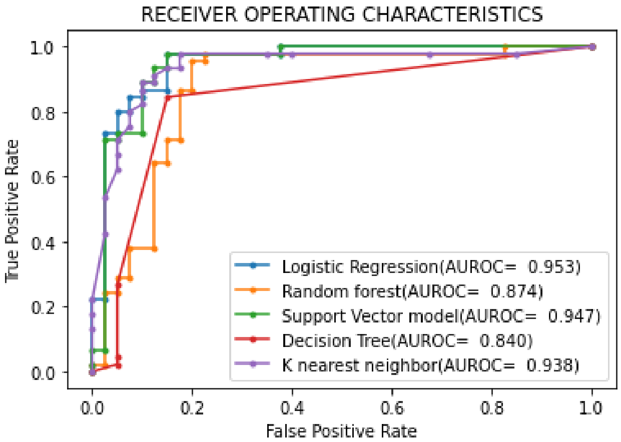


Fig. 7. Over-all ROC in detection of MA's and HE's.

According to this analysis, there are significantly fewer MAs and HEs in regions that are candidates than there are non-MAs and non-HEs. The ROC curve and corresponding AUC of MA detection results on stare database achieved by five classifiers (Decision tree, Random-forest, KNN, SVM and LR) are shown in Fig. 7. Fig. 8 describes the overall training accuracy of the six machine learning classifiers.

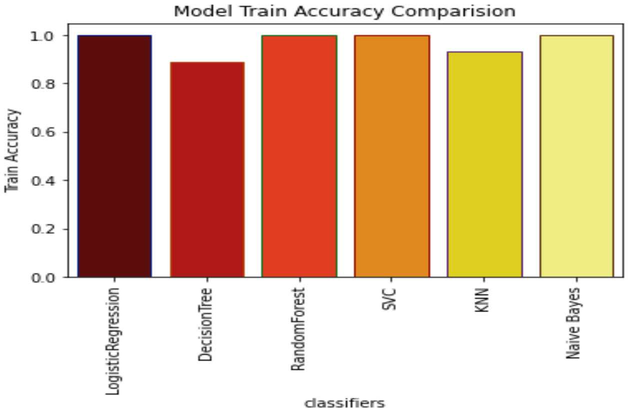


Fig. 8. Overall training accuracy of proposed model.

VI. Conclusion

The detection of MAs and HEs to identify the early stage of DR is displayed in this article. The suggested model consists of three main stages i.e., preprocessing, extraction of features and classification. In terms of accuracy, sensitivity, and specificity, respectively, the classifier logistic regression outperformed the other classification methods including DT, RFC, SVM, and KNN. However, because the preprocessing and extraction of feature steps are the only ones that control the whole system, the results are always a compromise between the necessary parameters. The suggested approach has been tested on databases such as STARE and evaluated using performance metrics as discussed in results. The Logistic Regression classifier (LRC) performance is good when compared with other classifiers with an accuracy of 94.6% in detection of MA's and 96.2% in detection of HE's. In future, research endeavors might involve integrating deep learning techniques into the machine learning algorithm and contrasting the outcomes with the ongoing study.

References

[1] R.Klein,B.E.K.Klein,S.E.Moss,Visualimpairmentindiabetes,Ophthalmology 91,1–9, 1994.

[2] Melville A, Richardson R, McIntosh A, O'Keeffe C, Mason J, Peters J, Hutchinson A. Complications of diabetes: screening for retinopathy and management of foot ulcers. Qual Health Care. 2000 Jun;9(2):137-41. doi: 10.1136/qhc.9.2.137.

[3] P.C.Ronald,T.K.Peng, A Textbook of Clinical Ophthalmology:A Practical Guide to Disorders of the Eyes and Their Management,3rd,World Scientific Publishing Company,Singapore,2003.

[4] Chaturvedi SS, Gupta K, Ninawe V, Prasad PS. Advances in computer-aided diagnosis of diabetic retinopathy. arXive-prints, 1909–09853 (2019). 1909.09853.

[5] Saha R, Chowdhury AR, Banerjee S. Diabetic retinopathy related lesions detection and classification using machinelearning technology. 2016;734–45.

[6] Seoud L, Hurtut T, Chelbi J, Cheriet F, Langlois JMP. Red lesion detection using dynamic shape features for diabeticretinopathy screening. IEEE Trans Med Imag. 2016;35(4):1116–26.

[7] Biyani RS, Patre BM. Algorithms for red lesion detection in diabetic retinopathy: a review. Biomed Pharmacother.2018;107:681–8.

[8] Srivastava R, Duan L, Wong DWK, Liu J, Wong TY. Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels. Comput Methods Progr Biomed. 2016;138:83–91.

[9] Zhou W, Wu C, Chen D, Yi Y, Du W. Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method. IEEE Access. 2017;5:2563–72.

[10] Wu B, Zhu W, Shi F, Zhu S, Chen X. Automatic detection of microaneurysms in retinal fundus images. Comput MedImaging Graph. 2017;55:106–12.

[11] Wang S, Tang HL, Turk LA, Hu Y, Sanei S, Saleh GM, Peto T. Localizing microaneurysms in fundus images through singular spectrum analysis. IEEE Trans Biomed Eng. 2017;64(5):990–1002.

[12] Adal KM, Sidibe D, Ali S, Chaum E, Karnowski TP, Meriaudeau F. Automated detection of microaneurysms usingscale-adapted blob analysis and semi-supervised learning. Comput Methods Progr Biomed. 2014;114(1):1–10.

[13] Derwin DJ, Selvi ST, Singh OJ. Secondary observer system for detection of microaneurysms in fundus images using texture descriptors. J Digit Imaging. 2019;1–9.

[14] Javidi M, Pourreza HR, Harati A. Vessel segmentation and microaneurysm detection using discriminative dictionary learning and sparse representation. Comput Methods Progr Biomed. 2017;139:93–108.

[15] M. Preethi. and R. Vanithamani., "Review of retinal blood vesseldetection methods for automated diagnosis of Diabetic Retinopathy,"IEEE-International Conference On Advances In Engineering,Science And Management (ICAESM -2012), Nagapattinam,Tamil Nadu, 2012,pp. 262-265.

[16] V. Saravanan, B. Venkatalakshmi and V. Rajendran, "Automated redlesion detection in diabetic retinopathy," 2013 IEEE Conference onInformation & Communication Technologies, Thuckalay, TamilNadu, India, 2013, pp. 236-239.

[17] M. Akter, M. S. Uddin, and M. H. Khan, "Morphology-basedExudates Detection from Color Fundus Images in Diabetic Retinopathy," International Conference on Electrical Engineering andInformation & Communication Technology (ICEEICT), 2014, pp. 1-4.

[18] A.K. Dixit and P. Prabhakar, "Hard Exudate Detection Using LinearBrightness Method", 4th International Conference on Recent Trendson Electronics, Information, Communication & Technology(RTEICT), 2019, pp. 980-984.

[19] G. G. Rajput and P. N. Patil, "Detection and Classification ofExudates Using K-Means Clustering in Color Retinal Images," 2014Fifth

International Conference on Signal and Image Processing,Bangalore, India, 2014, pp. 126-130.

[20] P. Kokare, "Wavelet based automatic exudates detection in diabeticretinopathy," 2017 International Conference on WirelessCommunications, Signal Processing and Networking (WiSPNET),Chennai, 2017, pp. 1022-1025.

[21] Auccahuasi, W. *et al.* Recognition of hard exudates using deep learning. *Procedia Comput. Sci.* **167**, 2343–2353 (2020).

[22] Prentasic, P. & Loncaric, S. Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion. *Comput. Methods Progr. Biomed.* **137**, 281–292 (2016).

[23] Sadek, I., Elawady, M. & Shabayek, A. E. R. Automatic classification of bright retinal lesions via deep network features. In *Computer Vision and Pattern Recognition* (2017).

[24] Abbasi-Sureshjani, S., Dashtbozorg, B., Romeny, B. M. H. & Fleuret, F. Boosted exudate segmentation in retinal images using residual nets. In Fetal, Infant and Ophthalmic Medical Image Analysis International Workshop, FIFI 2017 and 4th International Workshop OMIA 2017 Held in Conjunction with MICCAI 2017, Proceedings, 210–218. (Springer, 2017).

[25] Khojasteh, P. *et al.* Exudate detection in fundus images using deeply-learnable features. *Comput. Biol. Med.* **104**, 62–69 (2019).

[26] Joshi, S.; Karule, P. Mathematical morphology for microaneurysm detection in fundus images. Eur. J. Ophthalmol. 2020, 30, 1135–1142.

[27] Harangi, B.; Toth, J.; Hajdu, A. Fusion of Deep Convolutional Neural Networks for Microaneurysm Detection in Color Fundus Images. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018.

[28] Eftekhari, N.; Pourreza, H.-R.; Masoudi, M.; Ghiasi-Shirazi, K.; Saeedi, E. Microaneurysm detection in fundus images using a two-step convolutional neural network. Biomed. Eng. Online 2019, 18, 67.

[29] Xia, H.; Lan, Y.; Song, S.; Li, H. A multi-scale segmentation-to-classification network for tiny microaneurysm detection in fundus images. Knowl.-Based Syst. 2021, 226, 107140.

Gamers Intention Towards Purchasing Game Items in Virtual Community: Extending the Theory of Planned Behavior

Abbi Nizar Muhammad, Achmad Nizar Hidayanto
Magister of Information Technology
University of Indonesia
Jakarta, Indonesia

***Abstract*—Virtual communities serve as bustling marketplaces where gamers engage in transactions for in-game items, driving the digital economy's expansion. This research aims to illuminate the determinants of steering users' decisions within these online environments. Focusing on the constructs of Attitude, Subjective Norms, and Perceived Behavioral Control derived from the Theory of Planned Behavior (TPB), we investigate the factors shaping purchasing intentions. Employing structural equation modeling (SEM) on a robust dataset of 300 validated respondents, our analysis unveils insights into user motivations. Notably, the amalgamation of Attitude, Subjective Norms, and Perceived Behavioral Control explains 84% of the drivers guiding in-game item transactions within virtual communities. Our findings underscore the significance of certain attributes. Specifically, the perceived wisdom inherent in these transactions, the constructive influence of community discussions, and the ease of communication and negotiation channels within virtual realms emerge as pivotal determinants influencing user behavior. This study not only contributes to understanding user behavior in virtual spaces but also holds practical implications for scholars and industry stakeholders. By shedding light on these influential factors, this research informs strategies and interactions within virtual communities, offering valuable insights into the dynamics of the digital marketplace.**

Keywords—Theory of planned behavior; in-game items; purchasing intention; virtual community

I. Introduction

In the ever-evolving landscape of online gaming, virtual communities have emerged as dynamic and immersive spaces where players come together to share experiences, strategies, and a deep passion for their favorite games. Virtual communities often house bustling marketplaces where in-game items and assets are bought, sold, and traded amongst players. The allure of acquiring unique items or enhancing gaming experience has fueled the growth of the digital economy. According to Statista report, Global vide game revenue stands at USD 249.60 billion in 2023 and is projected to grow at a CAGR of 9.32% through 2028, reaching USD 389.70 billion [1].

Virtual communities take shape when a collective of individuals congregate to foster social connections through interactive communication within the digital realm [2]. The success and sustainability of an online community are contingent upon the willingness of its members to openly express their perspectives, engage in meaningful dialogues with fellow members, and make substantive contributions to the community's vitality by generating valuable content [3]. The ease and flexibility offered by this environment enhance communication among users, attracting gamers who frequent virtual communities for discussions and in-game item exchanges.

The quality of the virtual community significantly influences the perception of trust. Trust within an online environment plays a pivotal role in shaping users' intentions when it comes to transactions [4]. In the context of digital in-game items, trustworthiness holds particular significance, as these products are transmitted and utilized in a purely digital format. Building trust relies on user contributions within virtual communities, often through leveraging user reputation systems. The main purpose of user reputation systems is to establish trust between unknown parties [5].

Purchase intention in purchasing in-game items within virtual community can be analyzed on internal and external factors. In mobile games, the intention to purchase in-game items is driven by several factors such as progress in playing, competition, frequency of purchases and amount of spending [6]. Essential elements on internal factor of the purchase intentions of virtual in-games are enjoyment, skills, challenge, telepresence, and flow [7]. Previous research did not cover external factors, such as player behavior and motivation, which influenced purchase intentions. This study aims to explore these factors impacting players' behaviors and motivations in purchasing in-game items within virtual communities. Analyzing these aspects in purchase intention helps understand the driving forces behind a user's decision, providing valuable insights into their motivations within specific contexts.

This research explores the interesting phenomenon of game players' intentions regarding purchasing in-game items in virtual communities. It explores the complex interactions between psychological factors, social influences, and perceived control that guide individuals in their decisions to engage in these virtual markets. In doing so, it extends the renowned Theory of Planned Behavior to explain the motivations and intentions that drive these unique economic interactions.

The Theory of Planned Behavior (TPB) is an extension model of the theory of reasoned action and one of the most widespread models for social psychologists to predict behavioral intentions [8]. The theory of planned behavior (TPB) is one of the most widely researched frameworks for predicting behavioral intentions [9]. This theory investigates the factors that influence a person's behavior and how attitudes, subjective norms, and perceived behavioral control can influence a person's intention to engage in a particular action. By extending the Theory of Planned Behavior to the realm of virtual communities, this research seeks to uncover the unique variables that shape game players' intentions regarding the purchase of in-game items.

II. Theoretical Background

A. In-game Items

In-game items is virtual goods that can be defined as entities within the online gaming environment, which include characters, items, virtual currency, and tokens. Virtual goods refer to the subset of the virtual asset that can be mass-produced, bought, and sold live conventional consumer products, including items, characters, and currencies of games. Very often, three attributes of such virtual goods would drive players to purchase: functional, hedonic, and social [10].

In-game items are often associated with delivering hedonic value to users, prioritizing emotional gratification over practical utility. Hedonic products, in this context, tend to carry symbolic significance. Users invest in these items, receiving in return emotional satisfaction and pleasure as part of the benefits [11].

Gamers are inclined to acquire in-game items when these items enhance their gaming strength, add to their enjoyment, and enable them to showcase themselves to fellow players, especially when these items are reasonably priced. Consequently, within the realm of online gaming, a gamer's perception of an item's value significantly heightens the likelihood of them making a purchase.

B. Related Research

Guo and Yue developed a model that blends new constructs with established theories, drawing from frameworks such as the Theory of Planned Behavior (TPB), the Technology Acceptance Model (TAM), trust theory, and the Unified Theory of Acceptance and Use of Technology (UTAUT). Their research seeks to elucidate the factors behind consumers' decisions to purchase virtual items within virtual worlds [12]. The model highlights various components that bolster behavioral intentions, including trust, social influence, and perceived enjoyment. These elements are interconnected within the virtual community, specifically facilitating the trade of in-game items.

Ezlika conducted an examination of the determinants affecting players' purchase intentions regarding specific in-game virtual items. Additionally, the study explored the impact of the state of “flow" on these purchase intentions. The findings of this research unveiled that factors such as enjoyment, skill, challenge, and telepresence positively contribute to the experience of flow. Moreover, the study established a significant and positive association between the state of flow and players' purchase intentions [7]. Enjoyment, skill, challenge, and telepresence contribute significantly to the gaming experience, shaping gamers' purchasing behavior concerning in-game items. While gaming experience encapsulates an Attitude, it's imperative to quantify and assess these additional components for a comprehensive understanding.

In various fields, the Theory of Planned Behavior has been applied to analyze consumer intentions regarding product purchases in developing nations. Research conducted by Rambalak Yadav has demonstrated that Theory of Planned Behavior is effective in predicting consumers' intentions to purchase products [13]. Building upon prior research to discern the determinants of user purchase intention within virtual communities, the research employs the Theory of Planned Behavior. This theory helps identify several external factors, specifically centered around attitudes, subjective norms, and perceived behavioral control.

C. The Theory of Planned Behavior (TPB)

The Theory of Planned Behavior, initially developed by Ajzen in 1985, provides a well-established framework for understanding human behavior and decision-making across a spectrum of contexts, with intentions influenced by three main constructs [8].

1) Attitude: This construct reflects an individual's overall evaluation of the behavior, including the perceived outcomes and the subjective assessment of the behavior's desirability. In the context of buying in-game items, attitude encompasses players' beliefs about the consequences of such purchases, including how it enhances or affects their gaming experiences [13].

2) Subjective norms: Subjective norms refer to the perceived social pressure or influence from significant others, such as friends, family, and the gaming community, regarding the behavior in question. In the context of in-game item purchases, subjective norms encompass the influence of fellow gamers and the broader gaming culture on an individual's decision [13].

3) Perceived behavioral control: This construct pertains to the perceived ease or difficulty of performing the behavior and reflects the individual's perception of their ability to exercise control over it. In the context of virtual community transactions, perceived behavioral control includes factors like financial means, accessibility, and gaming expertise [13].

Extending the Theory of Planned Behavior to the realm of virtual community transactions is essential to unravel the complex motivations and intentions that underlie game player behavior.

D. Hyphothesis Development

This research aims to apply the Theory of Planned Behavior as an analytical framework to investigate and understand the intentions of game players regarding the purchase of in-game items within virtual communities.

We have identified certain limitations in the purchasing process in virtual communities, most notably the vulnerability of some virtual community's users to scam and fraud when acquiring products [14]. Scams and fraudulent activities continue to occur in virtual communities due to the lack of transactional features. Instead, we have identified the main advantages associated with purchasing in virtual communities, including aspects of simplicity, speed, and convenience. Additionally, users are incentivized to place trust in fellow community members, as forums offer an important space for discussion, access to reviews, informative content, and valuable feedback, all of which significantly influence their purchasing decisions [12]. With the limitations and benefits complementing each other, the existence of virtual communities persists. The discussion leads to following hypotheses:

H1. Game players' attitudes shaped by willingness purchase which has an impact on gaming experiences.

Player's attitude towards making in-game purchases could affect how they engage with the game.

H2. Subjective norms can encompass the influence of peers within the virtual community.

Virtual communities, especially gaming ones, peer influence plays a significant role. What others in the community do or say might shape an individual's beliefs about what is normal or acceptable behavior within that community.

H3. Perceived behavioral control includes consideration of financial resources, technical proficiency, and security concerns within the virtual community.

In virtual communities, various factors can affect how individuals navigate and participate. Financial resources might determine the extent to which someone can engage in certain activities or make purchases. Technical proficiency and security concerns also impact how comfortably and confidently someone can interact within that online space.

By exploring the interplay of attitudes, subjective norms, and perceived behavioral control, this study endeavors to contribute to a more comprehensive understanding of the motivations that guide players' purchase decisions within the virtual community, explains the complex relationships underlying the virtual economy of online games.

Having established the theoretical foundations that shape our understanding of purchase intention within virtual community. The following section delves into the methodology applied, explaining the systematic approach taken to empirically explore and validate the previously proposed hypotheses. The methodology section outlines the research design, data collection methods, and analytical framework used to investigate the complex dynamics in purchase intention within virtual community, bridging theory with practical implementation.

III. Research Methodology

This section describes the research methodology adopted for investigating game player intentions regarding the purchase of in-game items within virtual communities.

A. Research Instrument

This study employs a quantitative research design supported by the Theory of Planned Behavior (TPB) as the guiding framework. The research aims to explain the relationships between Theory of Planned Behavior constructs and game player intentions. Data collected through a structured survey questionnaire using a 5-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). The 5-point scale stands out for its simplicity, benefit for respondents due to its ease of comprehension and usage. In comparison to scales with more points, it offers a quicker and less demanding completion process.

To accomplish the objective of the study, all the constructs were adopted from the relevant literature. The present study showed the attitude towards virtual goods and the perceptions of peers' attitudes significantly bolster the inclination to engage in virtual goods procurement [15]. Constructing and measuring items were based on the scope of prior research that has empirically tested and enhanced the predictive capacity of the Theory of Planned Behavior [13]. The construct's significance in shaping behavioral intentions has been thoroughly tested. To align purchase intentions within virtual communities, constructs and measurement items were meticulously developed and tailored to fit the research specific context. Table I presents an overview of all the constructs referred to in this research.

TABLE I. CONSTRUCTS AND MEASURING ITEMS

Code	Constructs and Measuring Items
ATT	**Attitude**
ATT1	For me, buying in-game items via virtual communities is a good decision
ATT2	Purchasing in-game items from other users in virtual communities can enhance the gaming experience
ATT3	For me, buying in-game items via virtual communities is very fun
ATT4	For me, buying in-game items via virtual communities is a wise decision
ATT5	I believe that purchasing in-game items from other users in virtual communities is a worthwhile investment
ATT6	I feel satisfied when purchasing in-game items through virtual communities
SN	**Subjective Norms**
SN1	Community members and people I know in virtual communities encourage me to buy in-game items
SN2	Discussions in virtual communities have a positive impact on my decision to purchase in-game items
SN3	I trust the opinions and recommendations of other users in virtual communities regarding purchasing in-game items
PBC	**Perceived Behavioral Control**
PBC1	I have the virtual currency, time, and resources to make purchases of in-game items in virtual communities
PBC2	I can easily communicate and negotiate with other users in virtual communities regarding game item transactions
PBC3	I feel confident in my ability to purchase in-game items safely and securely in virtual communities
PI	**Purchase Intentions**
PI1	I intend to purchase in-game items from other users in virtual communities soon
PI2	I actively seek opportunities to purchase in-game items from fellow virtual community members
PI3	I plan to allocate part of my gaming budget to purchasing in-game items in virtual communities

B. Data Collection

Quantitative data collected through the distribution of survey questionnaires. The surveys administered online to a diverse sample of game players actively engaged in virtual communities. Active engagement within the virtual community defines gamers who consistently trade in-game items via forums or similar platforms. This segment actively participates in discussions and frequently engages in the exchange of in-game items. The questionnaire includes Likert-scale items designed to assess attitudes, subjective norms, perceived behavioral control, and intentions, according to the Theory of Planned Behavior framework.

Data pertaining to in-game items within virtual communities is predominantly collected from sources such as relevant discussion forums and other online community platforms [15]. Unexplainably, the pre-filtered data contained approximately 1,412 partially incomplete responses. Our investigation led us to the conclusion that this anomaly was a result of the survey software retaining incompletely filled questionnaires, despite being configured to exclusively accept fully completed ones. Fortunately, this software issue does not seem to have impacted the integrity of the collected data.

To address this issue, we employ Excel to filter the data before conducting our analysis using the Partial Least Squares Structural Equation Modeling (PLS SEM). The filtered dataset exclusively comprises questionnaires completed by respondents, ensuring a focused and relevant dataset for our analysis.

C. Data Analysis

Data analysis conducted using Partial Least Squares Structural Equation Modeling. This technique is suitable for examining complex relationships within the Theory of Planned Behavior framework and provides a robust method for assessing the structural model's fit to the data. The Partial Least Squares Structural Equation Modeling can be used for evaluation of the measurement model, evaluation of the structural model, goodness of fit estimation and hypothesis testing [16].

Partial Least Squares Structural Equation Modeling is a statistical methodology primarily employed for confirmatory purposes, validating hypotheses, and dissecting a structural theory that encompasses various phenomena within a specific environment. It serves as a powerful tool to test and confirm relationships between observed and latent variables, offering insights into the complex interplay among multiple factors within a given context or system. [17]. The stages in Partial Least Squares Structural Equation Modeling analysis encompass model specification, assessment of the outer (measurement) model, evaluation of the inner (structural) model, and ultimately, the examination of moderating variables to analyze their effects. This sequential process allows for a comprehensive understanding of relationships between constructs and potential moderating influences within the model under study [18].

Partial Least Squares Structural Equation Modeling used to validate the measurement model, ensuring that the survey items accurately capture the Theory of Planned Behavior constructs. This step involves assessing the reliability of factor loading (Cronbach’s Alpha), composite reliability, average variance extracted (AVE), convergent validity and discriminant validity were examined. Cronbach Alpha and Composite Reliability prove that the research model has good reliability. Cronbach alpha measures the internal consistency of a variable. The construct scores must be greater than 0.7 should be considered reliable [18]. Average variance extracted is a metric for determining convergent validity, Average variance extracted values is at least 0.5 and the square root of the Average variance extracted should be higher than correlation between one construct and the other constructs items [16].

The structural model examines the relationships between Theory of Planned Behavior constructs (attitude, subjective norms, and perceived behavioral control) and game player intentions. Partial Least Squares Structural Equation Modeling is particularly suited for research aiming to explore and predict relationships between latent variables, especially in situations where the underlying theory might be underdeveloped or relatively weak. It's an effective method for modeling complex relationships, making it valuable when theoretical frameworks are still evolving or when there's a need to explore and validate relationships between constructs that lack established theoretical foundations.

IV. Results and Discussion

Data was gathered from gamers in Indonesia who actively participate in trading within virtual communities through an online survey distribution. While a total of 1,412 responses were received from the distributed questionnaires, the final sample included 300 respondents who completed the survey in its entirety.

A. Demographics

Demographic characteristics of the respondents indicate that the largest portion, comprising 42%, falls within the 15 to 20 age group, while 35% belong to the age group below 15 and 23% above 20 age groups. Furthermore, a majority of the respondents, accounting for 50%, reported actively participating in trading within virtual communities for less than one year, with an additional 23% having between one to two years and 27% above two years of experience in game trading within virtual community.

B. Results

The study's findings were analyzed through the application of structural equation modeling, encompassing both the evaluation of the measurement model and the structural model analysis.

1) Measurement model assessment: We assess the measurement model's validity and reliability, which includes:

a) Validity analysis: Our validity analysis confirms the convergent validity, as all constructs exhibit outer loadings exceeding the established baseline of 0.7. Consequently, we have retained all constructs for further analysis. In the ATT construct, ATT4 stands out with the highest loading at 0.916, while in the SN construct, SN3 leads the way with an impressive 0.961. Within the PBC construct, PBC2 scores notably with a loading of 0.941, and in the PI construct, PI2 matches that value at 0.941.

Moreover, the Average Variance Extracted (AVE) values for each construct are all above 0.5, signifying the robustness and strong correlations within and across constructs. Specifically, ATT registers an AVE of 0.774, SN achieves 0.875, PBC attains 0.855, and PI maintains a solid 0.836. This observation highlights the substantial correlations within each construct and among different constructs.

In terms of discriminant validity, cross-loadings reveal that all factor correlations are below 0.8, signifying that adequate discriminant validity is maintained. Table II presents an overview of the validity analysis.

TABLE II. VALIDITY ANALYSIS

Code	Convergent Validity		Discriminant Validity			
	Outer Load-ing	AVE	Cross Loading			
			ATT	SN	PBC	PI
ATT1	0.830	0.774	0.830	0.695	0.705	0.656
ATT2	0.894		0.894	0.748	0.751	0.760
ATT3	0.903		0.903	0.815	0.791	0.757
ATT4	0.916		0.916	0.766	0.802	0.819
ATT5	0.848		0.848	0.755	0.778	0.840
ATT6	0.884		0.884	0.849	0.798	0.839
SN1	0.915	0.875	0.807	0.778	0.814	0.915
SN2	0.961		0.862	0.856	0.830	0.961
SN3	0.929		0.819	0.825	0.805	0.929
PBC1	0.915	0.855	0.815	0.915	0.805	0.810
PBC2	0.941		0.824	0.941	0.852	0.852
PBC3	0.919		0.797	0.919	0.813	0.770
PI1	0.895	0.836	0.783	0.817	0.895	0.758
PI2	0.941		0.863	0.845	0.941	0.822
PI3	0.907		0.758	0.780	0.907	0.814

b) Reliability analysis: In our Reliability Analysis, we assessed both Cronbach's Alpha (CA) and Composite Reliability (CR). In the field of social psychology research, CA and CR values above 0.7 are generally considered acceptable as a measure of internal consistency between items.

Cronbach's Alpha was employed to gauge the internal consistency of the items, and the study revealed that the values ranged from 0.902 to 0.941 within each construct. This range underscores the high degree of consistency within each construct.

In addition, construct reliability was assessed using Composite Reliability, with the study indicating values ranging from 0.904 to 0.943. These results clearly demonstrate that all values exceed the recommended threshold of 0.7 and are notably higher, underscoring the adequacy of the reliability of the constructs. Table III presents an overview of the reliability analysis.

TABLE III. RELIABILITY ANALYSIS

Code	Cronbach Alpha	Composite Reliability
ATT	0.941	0.943
SN	0.928	0.929
PBC	0.915	0.917
PI	0.902	0.904

2) *Structural model analysis:* The model analysis includes hypothesis testing, the assessment of path coefficients, and R-squared values.

a) Hypotheses testing: The analysis confirmed that the proposed theoretical framework successfully met the criteria for both reliability and validity. The results demonstrate that the proposed framework aligns well as a model fit. The visual representation illustrates the outcomes related to the postulated hypotheses.

Significantly, all the variables associated with the Theory of Planned Behavior - Attitude (value = 0.287, $p < 0.01$), confirming hypothesis H1 that customers are inclined to purchase items that enhance their gaming experience; Subjective Norm (value = 0.254, $p < 0.01$), substantiating hypothesis H2 that the decision to purchase in-game items is influenced by discussions and peer interactions within the virtual community; and Perceived Behavioral Control (value = 0.415, $p < 0.01$), supporting hypothesis H3 that factors like financial resources, technical proficiency, and security concerns play a role in the decision to purchase in-game items within the virtual community - demonstrated substantial relationships with gamers' intentions to acquire in-game items within a virtual community.

All the hypotheses have been confirmed by the findings, and the predicted values align with expectations. The graphical representation of the TPB model is displayed in Fig. 1.

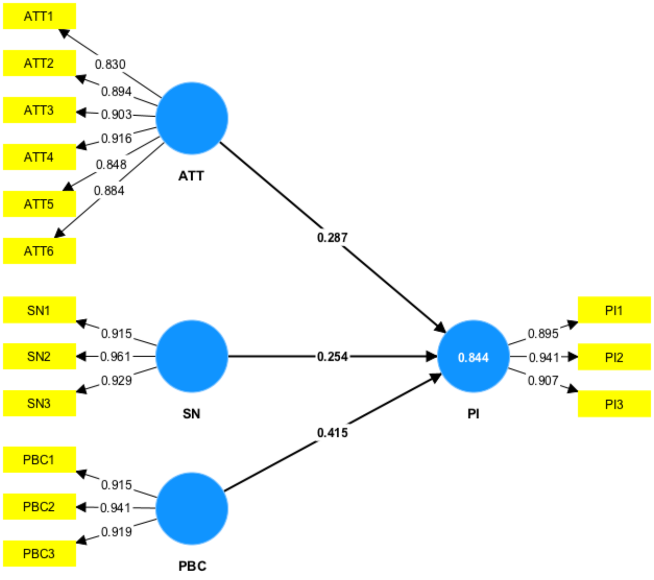


Fig. 1. Model Analysis.

b) Path coefficients: To support a research hypothesis, it's essential that the coefficient or direction of the variable relationship aligns with the hypothesized outcome, and the probability value (p-value) is less than 0.05 or 5%. In the study, the variables of the Theory of Planned Behavior (TBP) met this criterion, with Attitude (p-value = 0.005), Subjective Norm (p-value = 0.0001), and Perceived Behavioral Control (p-value = 0.013) all indicating values below the established threshold. This suggests that these variables are statistically significant and support the research hypothesis. Table IV presents an overview of the probability value.

TABLE IV. PROBABILITY VALUE

Code	P Value
ATT => PI	0.005
SN => PI	0.0001
PBC => PI	0.013

c) R-squared values: The R-Square value of 0.844 suggests that approximately 84.4% of the variation in the dependent variable (represented by Purchase Intention) can be explained by the independent variable included in the research model. The remaining variation, about 15.6%, is attributed to factors or independent variables that were not considered in this study.

C. Discussion

In exploring the implications of the relationships identified between the Theory of Planned Behavior constructs and gamer intentions within virtual community transactions. The discussion delves into the significance findings in the context of online gaming and virtual economies. Understanding the factors that influence purchase intentions in virtual communities has broader implications for consumer behavior in online environments.

The observed connections between attitude, subjective norms, perceived behavioral control, and purchase intentions resonate with the hypothesis and models, which underscores that the application of this framework in understanding gamer’s behavior in the context of virtual environment is very appropriate.

In earlier research within the same domain, it was found that the enjoyment of the game tended to decrease the willingness to acquire virtual goods. On the contrary, a positive attitude toward virtual goods and beliefs about peers' attitudes had a strong positive influence on the inclination to purchase such items [15]. Additionally, previous studies have indicated that internal factors like enjoyment, skill, challenge, and telepresence play a constructive role in establishing a significant connection between the state of "flow" and a player's intention to make in-game purchases [7]. While internal factors like enjoyment and challenge remain pivotal, external factors, specifically engagement within virtual communities and seamless communication channels, emerge as critical determinants shaping purchasing decisions.

This research extends the previous study by exploring influential external factors within virtual communities that drive purchase intentions for in-game items. The findings suggest that engagement within virtual communities has a positive impact on the decision to purchase in-game items within virtual community Furthermore, the ease of communication and negotiation with fellow users within the same virtual community emerges as a driving force behind purchase intentions.

D. Implications

These findings carry significant implications that could assist game developers and entrepreneurs in formulating effective strategies for the sale of in-game items within virtual communities. Based on these findings, it is advisable for entrepreneurs engaged in selling in-game items within virtual communities to offer comprehensive information to prospective buyers.

The content generated within these virtual communities, including discussions that give rise to opinions and recommendations, can foster a sense of trust among customers, ultimately influencing their decisions to make a purchase. Virtual communities should focus on fostering discussions and interactions among users. The positive impact of discussions, as shown in the study, suggests that strategies aimed at enhancing community engagement can be particularly effective.

Consistent with The Theory of Planned Behavior, it was observed that attitude, subjective norms, and perceived behavioral control emerged as notable and positively correlated predictors of the purchase of in-game items within virtual communities. The empirical results further indicated that subjective norms exhibited a more pronounced impact on the intention to purchase in-game items compared to attitude and perceived behavioral control. Lastly, attitude exerted a greater influence than perceived behavioral control, suggesting that purchasing in-game items in virtual communities is driven by considerations of simplicity, speed, and convenience for the customers.

V. CONCLUSION

This paper synthesizes the key findings that explain intricacies of gamer intentions within virtual communities regarding in-game item purchases. These insights bear significant implications for stakeholders in the gaming industry, laying the groundwork for enhancing user experiences, community engagement, and refining strategies for in-game items transactions.

The study's findings reveal that the intention to purchase in-game items within virtual communities is significantly influenced by three key factors: Attitude, Subjective Norms, and Perceived Behavioral Control. Together, these variables account for a substantial 84% of the reasons why users opt to purchase in-game items within forums or virtual communities.

Among these variables, Attitude numbers 4 (ATT4) within the Attitude construct stands out, suggesting that buying in-game items through virtual communities is seen as a prudent choice. Additionally, Subjective Norms numbers 2 (SN2) indicates that discussions in virtual communities positively impact the decision to purchase in-game items, while Perceived Behavioral Control numbers 2 (PBC2) highlights the ease of communication and negotiation with other users in virtual communities regarding game item transactions. These findings shed valuable light on the driving forces behind user behaviors in this context.

Moving forward, stakeholders in the gaming industry can leverage these insights to transform user experiences and optimize strategies. Developers can prioritize mechanisms that facilitate informed and transparent transactions, enriching user perceptions of these virtual communities as trustworthy marketplaces. Community managers, armed with an understanding of the positive influence of discussions, can foster engaging dialogues to augment user interactions and retention. Entrepreneurs can capitalize on the importance of seamless communication by enhancing platforms that facilitate easy and efficient exchanges among users.

Furthermore, this research represents a distinctive contribution to existing literature by delving deeper into the interplay of ATT4, SN2, and PBC2, offering specific insights into their roles within the context of in-game item transactions in virtual communities. This nuanced exploration expands the understanding of gamer intentions, providing granular insights into the multifaceted drivers shaping user behaviors.

Future research can explore how these factors and purchase intentions evolve over time. Longitudinal studies can provide a deeper understanding of the dynamics within virtual communities. Complementary to the Theory of Planned Behavior, incorporating additional variables and constructs may further enrich our comprehension of user behaviors in these environments, presenting promising paths for comprehensive investigations.

Acknowledgment

This work is funded by Hibah Riset Internal Faculty of Computer Science, Universitas Indonesia with contract number: NKB-002/UN2.F11.D/HKP.05.00/2023.

References

[1] Statista, "Video Games Worldwide," 2023.

[2] N. Luo, Y. Wang, M. Zhang, T. Niu and J. Tu, "Integrating community and e-commerce to build a trusted online second-hand platform: Based on the perspective of social capital," Technological Forecasting and Social Change, vol. 153, 2020.

[3] E. Akar, S. Mardikyan and T. Dalgic, "User Roles in Online Communities and Their Moderating Effect on Online Community Usage Intention: An Integrated Approach," International Journal of Human–Computer Interaction, vol. 35, no. 6, p. 495–509, 2019.

[4] K. Wei, Y. Li, Y. Zha and J. Ma, "Trust, risk and transaction intention in consumer-to-consumer e-marketplaces: An empirical comparison between buyers' and sellers' perspectives," Industrial Management & Data Systems, vol. 119, no. 2, 2019.

[5] M. J. A. Goncalves, R. H. Pereira and M. A. G. M. Coelho, "User Reputation on E-Commerce: Blockchain-Based Approaches," Journal of Cybersecurity and Privacy, vol. 2, no. 4, pp. 907-923, 2022.

[6] Y.-n. Seo, Y. Jung, J. Sng and J. Park, "Rational or Irrational Decision? Examination on Gamers' Intention to Purchase Probability-Type Items," Interacting with Computers, vol. 31, no. 6, pp. 603-641, 2019.

[7] E. M. Ghazali, "A study of player behavior and motivation to purchase Dota 2 virtual in game items," Kybernetes, vol. 52, no. 6, 2023.

[8] I. Ajzen, "From Intentions to Actions: A Theory of Planned Behavior.," Action Control, pp. 11-39, 1985.

[9] M. Soliman, "Extending the Theory of Planned Behavior to Predict Tourism Destination Revisit Intention," International Journal of Hospitality & Tourism Administration, vol. 22, no. 5, pp. 524-549, 2019.

[10] J. Cai, D. Y. Wohn and G. Freeman, "Who Purchases and Why?: Explaining Motivations for In-game Purchasing in the Online Survival Game Fortnite," in Proceedings of the Annual Symposium on Computer-Human Interaction in Play, New York, NY, United States, 2019.

[11] Yoo and J. Mee, "Perceived Value of Game Items and Purchase Intention," Indian Journal of Science and Technology, vol. 8, no. 19, 2015.

[12] Y. Guo and S. Barnes, "Why People Buy Virtual Items in Virtual Worlds with Real Money," Database for Advances in Information Systems, vol. 38, no. 4, pp. 69-76, 2007.

[13] R. Yadav, "Young consumers' intention towards buying green products in a developing nation: Extending the theory of planned behavior," Journal of Cleaner Production, vol. 135, pp. 732-739, 2016.

[14] R. Rachmadi, "Online Game Marketplace for Online Game Virtual Item Transaction," in 8th International Congress on Advanced Applied Informatics (IIAI-AAI), Toyama, Japan, 2019.

[15] J. Hamari, "Why do people buy virtual goods? Attitude toward virtual good purchases versus game enjoyment," International Journal of Information Management, vol. 35, no. 3, pp. 299-308, 2015.

[16] C. Q. Nguyen, A. M. T. Nguyen and l. B. Le, "Using partial least squares structural equation modeling (PLS-SEM) to assess the effects of entrepreneurial education on engineering students's entrepreneurial intention," Cogent Education, vol. 9, no. 1, 2022.

[17] B. M. Byrne, "Structural equation modelling with AMOS: Basic concepts, applications, and programming," New York: Routledge Taylor & Francis Group, 2010.

[18] J. F. Hair, H. G. T. M. R. C. S. M. M and K. O. Thiele, "Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods," Journal of the Academy of Marketing Science, vol. 45, no. 5, pp. 616-632, 2017.

Deep Learning-based License Plate Recognition in IoT Smart Parking Systems using YOLOv6 Algorithm

Ming Li, Li Zhang*
College of Computer and Control Engineering,
Northeast Forestry University, Haerbin 150040, Heilongjiang, China

***Abstract*—License plate recognition (LPR) is pivotal for the seamless operation of Internet of things (IoT) and smart parking systems, ensuring the swift and effective identification and management of vehicles. Recent research has concentrated on refining LPR methods through deep learning approaches, proposing diverse strategies to enhance accuracy and reduce computation costs. This work tackles these challenges by introducing an innovative method rooted in the YOLOv6 algorithm. Leveraging a tailored dataset for model generation, the study employs rigorous methodologies involving validation, testing, and training. The resultant model demonstrates marked improvements in license plate recognition capabilities, surpassing the performance of existing methods. This breakthrough bears significant implications for advancing IoT smart parking systems, promising heightened reliability and efficiency in vehicle identification and management. Thorough experimental results and performance evaluations validate the efficacy of the proposed YOLOv6-based method. In-depth discussions and comparisons with state-of-the-art methods in the field lead to the conclusion that the introduced approach not only elevates accuracy but also enhances overall efficiency in license plate recognition for smart parking systems, thereby providing valuable contributions to the domain.**

Keywords*—Internet of things; deep learning; smart parking system; license plate recognition; YOLOv6

I. Introduction

With the rapid advancements in technology, the concept of smart cities has gained considerable attention in recent years. Smart cities leverage the power of the Internet of Things (IoT) to create intelligent, interconnected urban environments that enhance the quality of life for their residents [1, 2]. One of the key areas of focus in smart city development is smart parking management, which aims to address the persistent challenges associated with parking in urban areas [3, 4]. In this context, video-based technologies have emerged as a promising solution, offering real-time monitoring, efficient utilization of parking spaces, and improved enforcement capabilities.

The importance of video-based technologies in IoT smart parking systems cannot be overstated [5]. Traditional parking management systems often rely on physical sensors or manual monitoring, which can be cumbersome and time-consuming. Video-based technologies, on the other hand, utilize cameras and computer vision algorithms to capture and analyze visual data, enabling real-time tracking and intelligent decision-making [6, 7]. These technologies provide valuable insights into parking occupancy, duration, and violation detection, leading to enhanced efficiency and improved parking experiences for drivers [8]. Moreover, video analysis can be seamlessly integrated into existing IoT platforms, facilitating data-driven decision-making and enabling effective management of parking spaces in smart cities.

License plate recognition (LPR) plays a vital role in smart parking management systems. By leveraging computer vision techniques, LPR systems automatically capture, interpret, and recognize license plate information from video streams or images [9]. Integrating LPR into smart parking systems enables automated entry and exit control, efficient payment processing, and enhanced enforcement capabilities [10]. This technology eliminates the need for manual checks or physical tickets, streamlining the parking process and reducing human intervention. License plate recognition serves as a fundamental component in smart parking systems, facilitating seamless and secure parking operations within smart cities.

Various methods have been explored for license plate recognition in the past, with recent research focusing heavily on deep learning-based approaches [10]. Deep learning has garnered significant attention due to its ability to learn complex patterns and features from large-scale datasets [11-13]. Compared to traditional methods that rely on handcrafted features and rule-based algorithms, deep learning-based methods offer superior accuracy and robustness in license plate recognition tasks [14]. The availability of large labeled datasets and advancements in computational power have further fueled the adoption of deep learning in this domain, attracting researchers to explore innovative approaches and architectures.

Despite the progress made in deep learning-based license plate recognition systems, some limitations and research gaps still need to be addressed. One of the critical challenges lies in achieving a balance between computational cost and accuracy rate. Real-time processing of video streams requires low computational overhead while maintaining high accuracy in license plate recognition. The existing literature offers insights into these limitations, motivating the need for further research to develop lightweight deep learning and YOLO-based (You Only Look Once) algorithms. These approaches offer the potential to address the identified research gap by achieving a good trade-off between computational efficiency and recognition accuracy.

This study addresses the research problem of improving license plate recognition (LPR) in the context of internet of things (IoT) and smart parking systems. The existing challenges involve the necessity for more accurate and efficient vehicle identification and management. The research questions focus on evaluating the effectiveness of the YOLOv6 algorithm in overcoming these challenges, particularly in terms of enhancing accuracy and reducing computation costs. The research objectives aim to assess and optimize LPR using the YOLOv6 algorithm, with a specific emphasis on evaluating its accuracy in license plate identification, analyzing computational efficiency to manage costs, and contributing to the overall advancement of IoT and smart parking systems.

The main research contributions of this study are listed as follows:

1) Preparing a custom dataset and the application of YOLO-based algorithms and lightweight deep learning models to enhance license plate recognition performance in intelligent parking management systems.

2) Developing an efficient deep learning-based method for LPR to serve low computational cost and high accuracy rate requirements.

3) Conducting comprehensive experimental evaluations and performance analysis, highlighting the effectiveness and efficiency of the proposed methods.

The rest of the paper is structured as follows: In Section II, a review of previous studies is presented. Section III presents the methodology. Performance analysis is discussed in Section IV. The paper is finally concluded in Section V.

II. Review of Previous Studies

Loong et al. [15] presented a machine vision-based smart parking system utilizing the Internet of Things (IoT). The system aims to provide efficient parking management using camera-based image processing techniques to detect and monitor parking spaces in real-time. It utilizes IoT technologies to enable seamless communication between parking sensors, cameras, and a central server. The system allows users to access parking availability information through a mobile application, thereby reducing the time spent searching for parking spaces. However, the limitation of this paper is that it focuses primarily on the technical aspects of the system and does not extensively discuss the implementation challenges or potential scalability issues that may arise when deploying the system in real-world scenarios. Further research could investigate the practical implications and limitations of implementing the proposed smart parking system.

In study [16], a smart parking application is implemented that utilizes the Internet of Things (IoT) and machine learning algorithms. The system aims to optimize parking space utilization by monitoring and managing parking availability in real-time. It employs IoT technologies to collect data from parking sensors and employs machine learning algorithms to predict parking space occupancy. The system offers users a mobile application to access parking information and reserve parking spots. However, a limitation of this paper is that it lacks a comprehensive evaluation of the system's performance, scalability, and robustness in real-world scenarios. Future research could focus on assessing the system's effectiveness in different environments and under varying conditions to address potential limitations and enhance its practical implementation.

Kabir et al. [17] present an IoT-based intelligent parking system that focuses on utilizing unutilized parking areas. The system employs real-time monitoring utilizing mobile and web applications to provide users with parking availability information and facilitate efficient parking space utilization. It utilizes IoT technologies to collect data from parking sensors and employs intelligent algorithms for real-time monitoring. The system offers a user-friendly interface for accessing parking information and making reservations. However, a limitation of this paper is the lack of empirical validation or field testing of the proposed system in real-world parking environments. Further research could evaluate the system's performance, reliability, and scalability to address potential limitations and validate its practical effectiveness in different parking scenarios.

In [18], a smart parking system was implemented that utilizes image processing techniques. The system aims to optimize parking space management using cameras to detect and analyze parking space occupancy. It provides real-time parking availability information to users through a user-friendly interface. However, a limitation of this paper is the lack of discussion regarding the system's performance in challenging lighting conditions or crowded parking scenarios, which may affect the accuracy of the image processing algorithms. Future research could address these limitations by conducting thorough testing and evaluation of the system's performance in various real-world environments to ensure its reliability and effectiveness.

Deep learning algorithms for automatic license plate recognition (ALPR) were implemented by Silpa [19]. The study aims to develop an ALPR system that utilizes deep learning techniques to recognize and detect license plates in real accurately. The system demonstrates promising results in terms of accuracy and efficiency. However, a limitation of this paper is the absence of a comprehensive analysis of the system's performance under challenging conditions, such as variations in lighting, weather, or license plate designs. Future research could address these limitations by conducting extensive testing and evaluation of the system's robustness and reliability in various real-world scenarios to ensure its practical applicability.

The collective findings from previous studies on IoT-based smart parking systems reveal a common research challenge: achieving high accuracy in real-time parking space detection while minimizing computation costs. Despite notable advancements in parking management, lack of comprehensive performance evaluations, and absence of empirical validation in real-world scenarios. Addressing these challenges, future research should focus on developing smart parking systems that strike a balance between accuracy and computation costs. Thorough empirical validation and comprehensive evaluations are crucial to advancing the field and facilitating practical implementation in diverse urban environments.

III. Methodology

This section discusses the methodology of this study. In the methodology, we describe the structure of the suggested method. In the proposed method, the steps consisted of Data Collection and Preparation, Model Architecture Design, Dataset Preprocessing, Model Training, Model validation, and Model Testing. Fig. 1 illustrates the structure of the suggested method.

As shown in Fig. 1, the proposed method follows a structured approach consisting of several key steps. First, the Data Collection and Preparation phase involves gathering a diverse dataset of license plate images captured from real-world parking scenarios and annotating them with corresponding labels. Next, the Model Architecture Design step focuses on selecting a suitable deep-learning architecture and configuring its parameters to design an effective license plate recognition model. Following that, the dataset preprocessing step involves applying necessary preprocessing techniques to normalize the license plate images and augment the training dataset for improved model generalization. Subsequently, the Model Training phase utilizes the annotated dataset to train the deep learning model, optimizing a chosen loss function and fine-tuning hyperparameters. The trained model is then subjected to Model Validation, where its performance is evaluated on a validation dataset to ensure proper generalization. Finally, the Model Testing step involves applying the trained model to a separate testing dataset to assess its real-world performance, including accuracy, detection rate, and recognition speed. This structured approach ensures a comprehensive development and evaluation process for the proposed method in license plate recognition within smart parking management systems—the details of the steps are discussed in the following sections.

A. Data Collection and Preparation

In the data collection and preparation step, a diverse dataset of license plate images is collected from real-world parking scenarios and internet resources. This dataset aims to capture the variability and complexity encountered in actual parking environments. It includes images of license plates from different vehicle types, varying lighting conditions, different angles, and potential occlusions. By collecting a diverse dataset, the proposed method ensures that the trained model can handle the challenges faced in real-world scenarios.

Once the dataset is collected, the next action is to annotate it with corresponding labels and bounding boxes surrounding the license plates. This annotation process involves manually marking the regions of interest containing the license plates in each image and assigning the correct labels to them. The labels typically include alphanumeric characters present on the license plates. This annotation step is crucial as it provides ground truth information that is essential for training and evaluating the model accurately.

B. Model Architecture Design

For the design of our deep learning model for license plate recognition, we have chosen YOLOv6 as our core model. YOLO (You Only Look Once) is a popular object detection framework known for its real-time performance and accuracy. YOLOv6 is an evolution of the YOLO series, incorporating improvements and optimizations to enhance its detection capabilities.

YOLOv6 serves as an excellent starting point for our license plate recognition model due to its pre-trained weights and strong performance on object detection tasks. Pre-trained models are trained on large-scale datasets such as Common Objects in Context (COCO), which contain various object classes, including vehicles and license plates. Leveraging a pre-trained model like YOLOv6 allows our license plate recognition model to benefit from the knowledge gained during the pre-training phase, enabling faster convergence and better generalization.

With YOLOv6 as our core model, we can leverage its architecture, which is based on a deep neural network with convolutional layers, to localize and detect license plates within images. The YOLOv6 architecture employs a series of convolutional layers followed by fully linked layers to learn features at different scales, enabling it to detect objects with different sizes and aspect ratios accurately. By adopting the YOLOv6 model to our license plate recognition task, we can leverage its inherent object detection capabilities to identify and localize license plates within parking images. The pre-trained weights of YOLOv6 serve as a starting point for our model, allowing us to fine-tune and specialize the network for license plate recognition specifically. This approach combines the advantages of transfer learning, where pre-trained knowledge is utilized, with the flexibility to adapt the network to the requirements of license plate recognition in smart parking management systems.

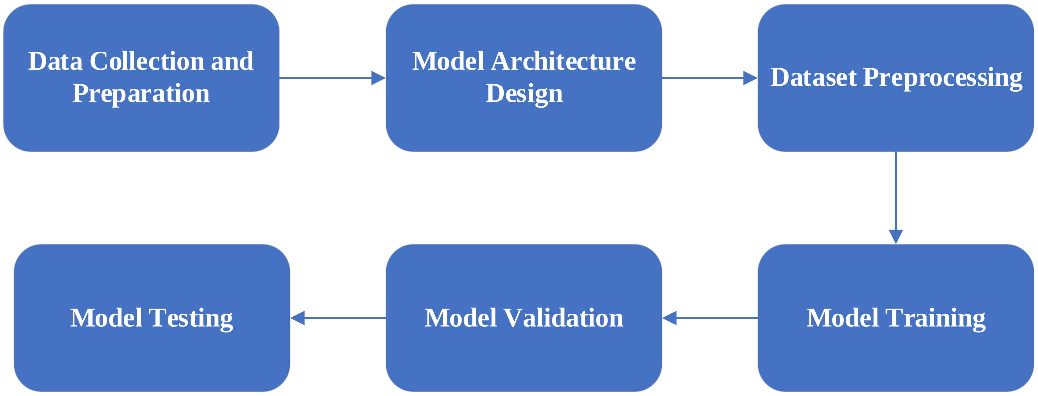


Fig. 1. The suggested method's structure.

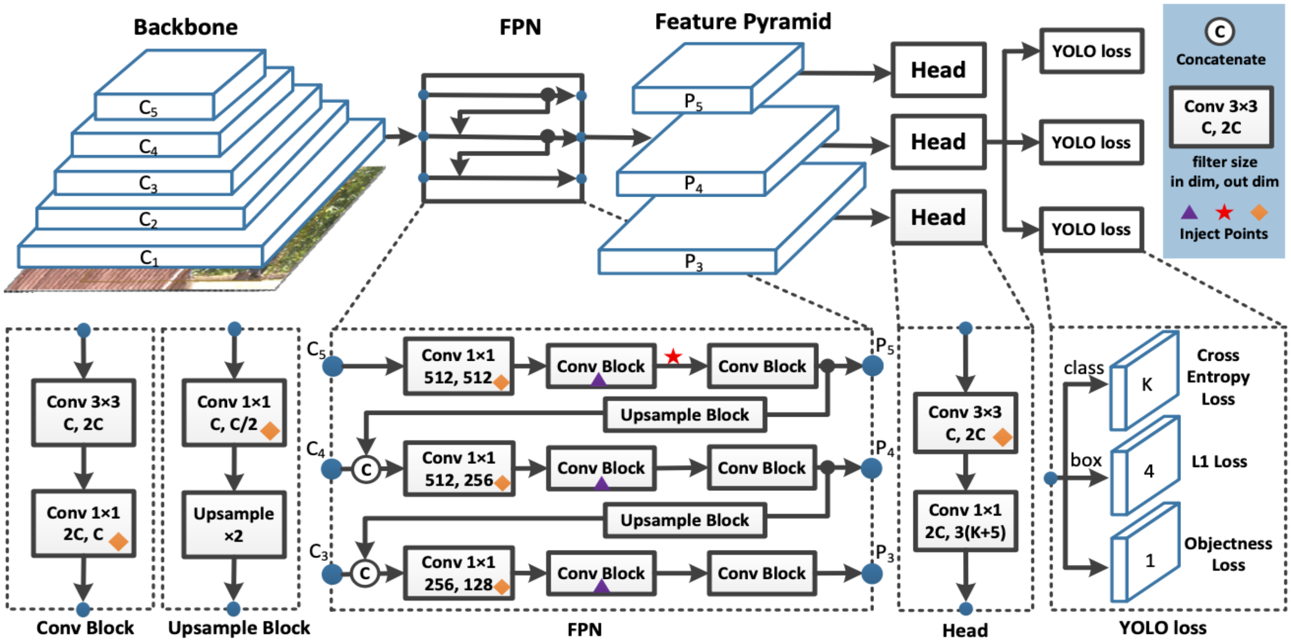


Fig. 2. Structure of YOLOv6 network [20].

The structure of the YOLOv6 network consists of a backbone, neck, and head for object detection. Fig. 2 illustrates the structure of the YOLOv6 network. The details of each are discussed in the following section.

1) The backbone architecture of YOLOv6 network: The backbone architecture of the YOLOv6 network serves as the foundation for its object detection capabilities. It incorporates a deep neural network with a series of convolutional layers to extract meaningful features from input images [21]. The backbone architecture of YOLOv6 follows a hierarchical structure comprising multiple residual blocks with different filter sizes. These residual blocks enable the network to learn and capture features at various spatial scales, allowing for the accurate detection of objects of different sizes and aspect ratios. Additionally, the backbone architecture incorporates skip connections that facilitate the flow of information across different layers, enabling the network to capture both low-level and high-level features. By leveraging this hierarchical backbone architecture, YOLOv6 can effectively analyze input images and generate precise bounding box predictions for objects, including license plates, with high accuracy and efficiency. Fig. 3 shows the structure of the backbone in the YOLOv6 network.

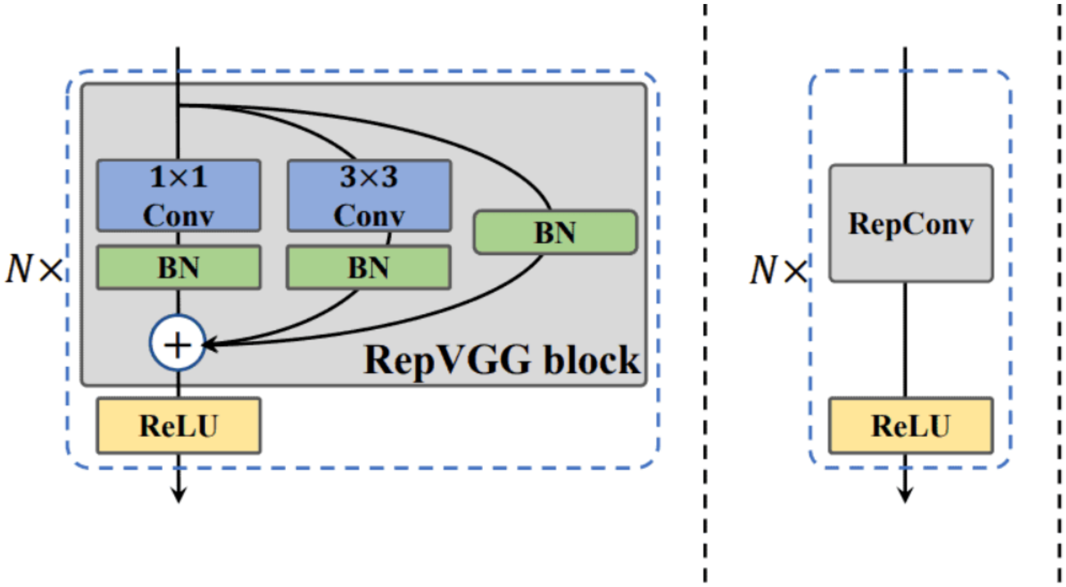


Fig. 3. The structure of the backbone in the YOLOv6 network [22].

2) The neck architecture of YOLOv6 network: The neck architecture of the YOLOv6 network complements the backbone architecture by refining and fusing features for more accurate object detection [21]. It incorporates spatial pyramid pooling modules to capture contextual information at multiple scales and feature fusion layers to combine features from different levels. This integration of features enhances the network's ability to detect objects of varying sizes and improves its overall precision and accuracy. By leveraging the neck architecture, YOLOv6 can effectively integrate global and local contextual cues, resulting in more robust and informed object detection predictions, including accurate identification of license plates.

3) The detection head of the YOLOv6 network: The detection head of the YOLOv6 network is responsible for generating precise bounding box predictions and class probabilities for detected objects, including license plates [23]. It takes the fused features from the neck architecture as well as processes them to identify and localize objects within the input image [21]. The detection head consists of a set of convolutional layers followed by anchor boxes, which serve as prior knowledge about the expected sizes and shapes of objects. Fig. 4 demonstrates the structure of the head in the YOLOv6 network.

As shown in Fig. 4, the convolution layers aid in capturing more detailed and discriminative information for accurate object detection. The anchor boxes, defined with different aspect ratios and scales, are used to predict bounding boxes for potential objects at various positions and sizes. The detection head employs a combination of objectness scores, class probabilities, and bounding box coordinates to generate the final predictions. Objectness scores indicate the presence of an object within a bounding box, while class probabilities specify the likelihood of the object belonging to a specific class, such as a license plate. The predicted bounding box coordinates provide precise localization of the detected object.

4) Analysis of YOLv6 performance: This section presents the analysis of the YOLOv6 performance. This analysis intends to report the efficiency of this algorithm used in object detection. Fig. 5 illustrates a graph for this analysis [21]. The graph represents a comparison of different YOLO models, namely YOLOv5, YOLOv6, YOLOv7, YOLOX, and PP-YOLOE, in terms of average precision (AP) with respect to latency. The X-axis represents latency, which refers to the time taken for the model to process input, and the Y-axis represents average precision (AP), a metric that measures the accuracy of the model's predictions.

As shown in Fig. 5, by analyzing the graph, we can observe the performance of each model in terms of AP at different latency levels. The comparison allows us to evaluate how well the models balance accuracy and speed in object detection tasks. Based on the graph, it is evident that YOLOv6 achieves better results compared to the other models across various latency levels. The higher AP values indicate that YOLOv6 provides more accurate object detection compared to YOLOv5, YOLOv7, YOLOX, and PP-YOLOE. The superiority of YOLOv6 can be attributed to several factors, such as architectural improvements, optimization techniques, or the incorporation of novel features or modules. To further justify why YOLOv6 outperforms other models, it would be necessary to refer to specific technical details and research papers related to YOLOv6. These details could include architectural enhancements, algorithmic advancements, or optimizations implemented in YOLOv6 that contribute to its improved accuracy while maintaining acceptable levels of latency.

This study justifies the selection of YOLO (You Only Look Once) as the base detection method for license plate recognition, highlighting its efficiency in real-time object detection with a balance between accuracy and speed. YOLO's unique single-pass approach and grid-based analysis set it apart from other methods. The research further strengthens this choice by comparing various YOLO-based versions, presenting experimental results that affirm YOLOv6 as superior. The comparisons consider accuracy, computation costs, and overall performance, providing empirical evidence that establishes YOLOv6 as the most effective and efficient version for license plate recognition based on conducted experiments results.

C. Dataset Preprocessing

In the dataset preprocessing phase, the license plate images undergo the necessary techniques to normalize them and augment the training dataset. In this study, normalization involves resizing, cropping, and color space conversion to standardize the images, and augmentation techniques such as rotation, scaling, and noise addition are applied to increase dataset diversity. These preprocessing steps ensure consistent image sizes, isolate the license plate region, and enhance the dataset's variability. By employing these techniques, the model becomes more robust, able to manage various scenarios, and generalizes better for accurate license plate recognition in smart parking management systems.

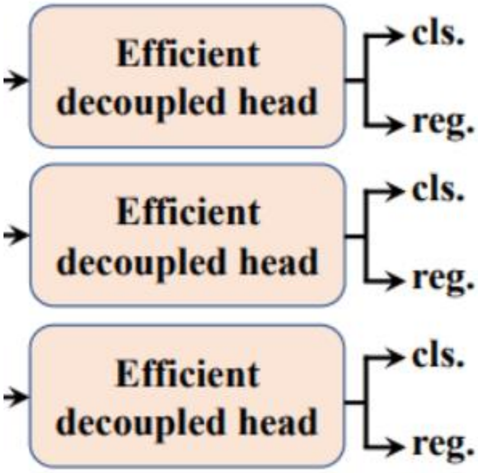


Fig. 4. Structure of detection head of YOLOv6 Network [22].

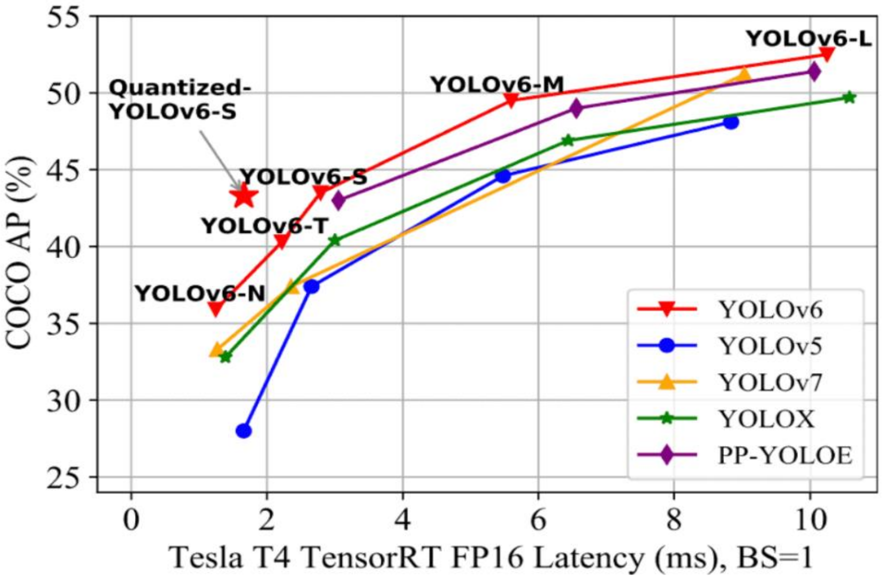


Fig. 5. Comparison of YOLO-based algorithms [21].

D. Model Training

We use transfer learning for YOLOv6 model training. To generate the YOLOv6 model for license plate recognition using transfer learning, the following steps are taken in the Model Training phase:

1) The deep learning model with the selected YOLOv6 architecture is initialized. The pre-trained weights from the base YOLOv6 model serve as the starting point for the transfer learning process. These weights contain valuable knowledge gained from pre-training on large-scale datasets, which aids in faster convergence and better generalization.

2) The initialized model is trained using the annotated license plate dataset. The training process involves feeding the license plate images as input to the model and optimizing a suitable loss function, such as categorical cross-entropy or mean squared error. The loss function measures the discrepancy between the forecasted outputs as well as the ground truth labels, driving the model to minimize the error and improve its performance.

3) Fine-tuning the model is a critical step to enhance its performance further. Hyperparameters, including the learning rate, batch size, and optimizer choice (such as Adam or SGD), are adjusted to achieve better convergence and generalization. Fine-tuning allows the model to adapt to the specific characteristics of the license plate recognition task and improves its ability to detect and recognize license plates accurately.

By following these steps, the pre-trained YOLOv6 model is effectively utilized as a starting point for training the license plate recognition model. Through transfer learning and fine-tuning, the model is optimized to leverage the knowledge from the pre-trained weights, adapt to the license plate recognition task, and achieve improved accuracy and performance in identifying and localizing license plates within smart parking management systems.

E. Model Validation

The YOLOv6 model for license plate recognition can be validated through evaluation on a validation dataset and the utilization of metrics such as accuracy, precision, recall, and F1-score. By assessing the model's performance on unseen data, researchers can detect overfitting or underfitting issues and make necessary adjustments. Metrics like accuracy, precision, recall, and F1-score provide quantitative measures of the model's license plate recognition capability, including overall correctness and avoidance of false positives and false negatives. Validating the YOLOv6 model using these approaches ensures its effectiveness and reliability in accurately identifying and localizing license plates in smart parking management systems.

F. Model Testing

After generating the YOLOv6 model and validation process, it is crucial to test and deploy the model effectively. Testing involves evaluating the model's performance using a diverse test dataset, comparing its predictions with ground truth labels, and calculating performance metrics such as accuracy, precision, recall, and F1-score. This testing phase ensures that the model performs well on unseen data and provides reliable license plate recognition results. Once the model has been thoroughly tested and meets the desired performance criteria, it can be deployed in real-world applications. Deployment involves integrating the model into the smart parking management system, optimizing its computational requirements for real-time performance, and considering factors such as handling input data streams, integrating with existing infrastructure, and creating an end-to-end pipeline for license plate recognition. Continuous monitoring and maintenance are essential post-deployment to monitor the model's performance, collect feedback, and address any issues that arise, ensuring consistent and reliable license plate recognition in real-world scenarios.

IV. Results and Discussions

Analyzing the generated YOLOv6 model for license plate recognition involves evaluating its performance using metrics such as precision, recall, and F1-score.

- Precision measures the proportion of correctly identified license plates among all the predicted license plates. It quantifies the model's ability to avoid false positives, which are instances where the model incorrectly identifies non-license plate regions as license plates. A higher precision shows that the model has a lower rate of false positives, resulting in more accurate and reliable license plate recognition.
- Recall, on the other hand, measures the proportion of correctly identified license plates among all the actual license plates present in the dataset. It assesses the model's ability to avoid false negatives, which occur when the model fails to identify actual license plates. A higher recall indicates that the model has a lower rate of false negatives, meaning it can effectively identify and capture most of the license plates present in the dataset.
- F1-score is a metric that combines precision and recall into a single value, providing a balanced measure of the model's overall performance. It is the harmonic mean of precision and recall and takes into account both false positives and false negatives. A higher F1 score indicates a better balance between precision and recall, signifying a model that can accurately identify license plates while minimizing both types of errors.

The results of precision curves, precision-recall curve, F1-score curve, and recall curve are shown in Fig. 5, 6, 7, and 8.

As shown in Fig. 6, the link between confidence and precision rate is graphically represented by the precision curve. The X-axis displays confidence levels, which indicate the model's certainty in its predictions, while the Y-axis displays the precision rate, which measures the proportion of correctly identified license plates among all the predicted license plates at different confidence thresholds. In the precision curve, as the confidence threshold increases, the precision rate tends to improve. This means that when the model is more confident in its predictions, it is more likely to correctly identify license plates, resulting in a higher precision rate. A high precision rate indicates that the generated model can achieve accurate results with a lower rate of false positives.

As demonstrated in Fig. 7, the X-axis displays confidence levels, indicating the model's certainty in its predictions, while the Y-axis displays the recall rate, which measures the proportion of correctly identified license plates among all the actual license plates at different confidence thresholds.

When analyzing the recall curve, as the confidence threshold increases, the recall rate typically decreases. This means that as the model becomes more conservative in its predictions, it may miss some actual license plates, resulting in a lower recall rate. However, a high recall rate indicates that the generated model can effectively identify a larger percentage of license plates in the dataset, minimizing false negatives.

Fig. 8 presents the Precision-Recall curve of the generated model. In this graph, the X-axis represents recall, which measures the proportion of correctly identified license plates among all the actual license plates, while the Y-axis displays the precision rate, which quantifies the proportion of correctly identified license plates among all the predicted license plates.

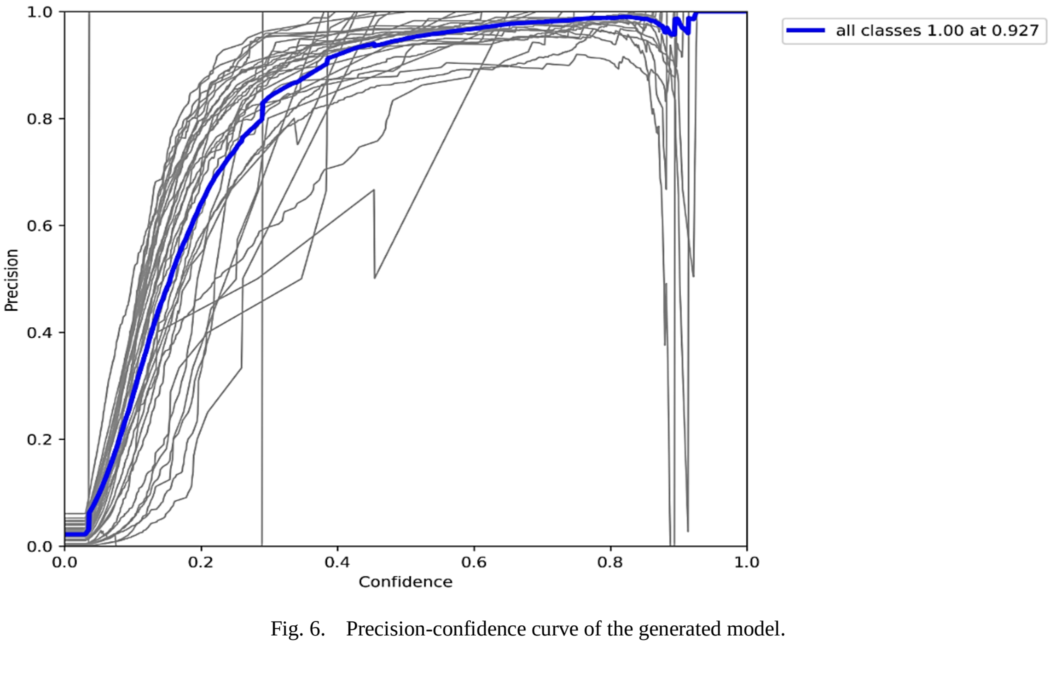


Fig. 6. Precision-confidence curve of the generated model.

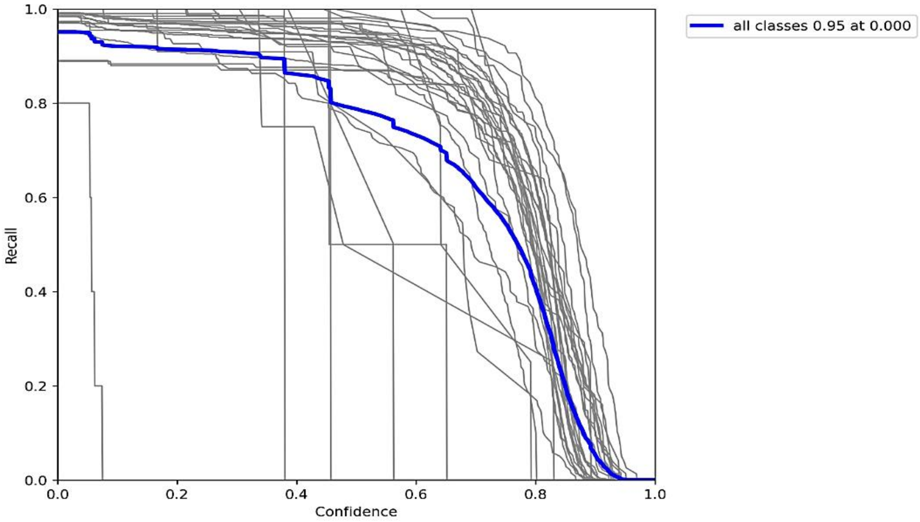


Fig. 7. Recall-confidence curve of the generated model.

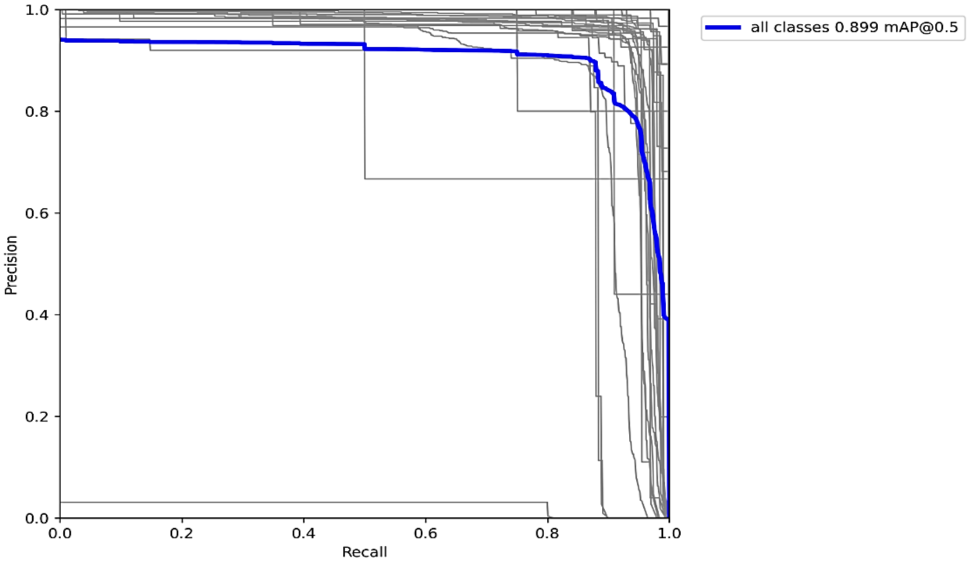


Fig. 8. The precision-recall curve of the generated model.

The precision-recall curve illustrates how changes in the recall threshold impact the precision rate and vice versa. As the recall threshold rises, the model becomes more inclusive and captures a higher percentage of actual license plates, resulting in an increase in recall. However, this may lead to more false positives, causing the precision rate to decrease. On the other hand, raising the precision threshold makes the model more conservative, reducing false positives but potentially missing some actual license plates, which lowers the recall rate. As experimental results show, analyzing the precision-recall rates obtained from the curve provides valuable insights into the accuracy and performance of the generated model in license plate recognition tasks.

Fig. 9 illustrates the F1-confidence curve, a graphical representation showcasing the relationship between confidence and F1 values. In this curve, the X-axis displays confidence levels, indicating the model's certainty in its predictions, while the Y-axis represents the F1 values, which is the harmonic mean of precision and recall.

The way that modifies how the confidence threshold impacts the F1 score is seen by the F1-confidence curve. As the confidence threshold increases, the model becomes more conservative in its predictions, resulting in a higher precision but potentially lower recall. Conversely, lowering the confidence threshold leads to a higher recall but may introduce more false positives, affecting precision. The F1 score strikes a balance between precision and recall, providing a comprehensive evaluation of the model's performance.

By analyzing the F1-confidence curve, we indicate that the generated model that achieves a high F1 score at an optimal confidence threshold indicates its capacity to precisely recognize license plates while minimizing false positives and false negatives.

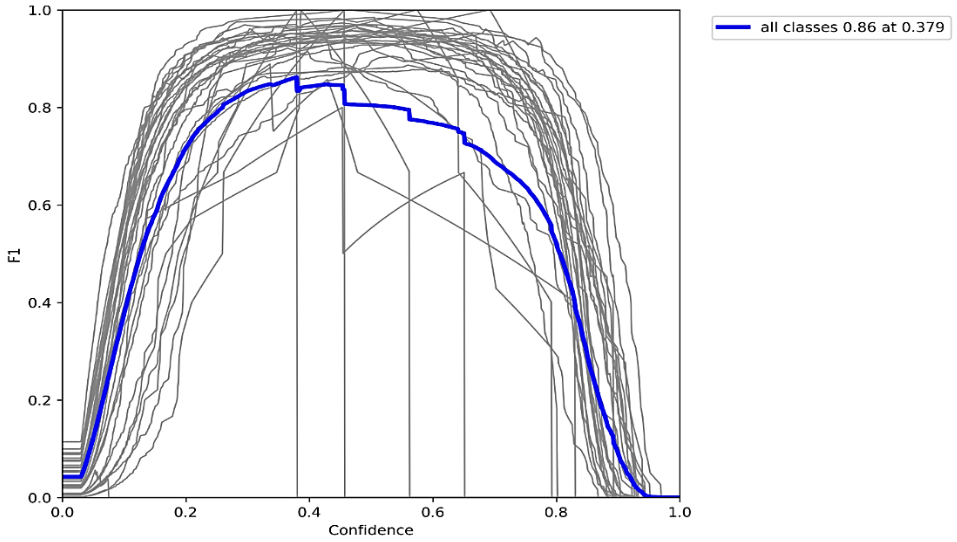


Fig. 9. F1-confidence curve of the generated model.

As reported results, the YOLOv6 achieves superior results in terms of high accuracy and low computation cost through a combination of innovative design features and improvements over its predecessors. The model's unique architecture, characterized by a single-pass approach and grid-based analysis, significantly contributes to its efficiency in real-time object detection. By efficiently dividing the input image into a grid and processing it in a single pass through the neural network, YOLOv6 minimizes computation costs while ensuring accurate detection. The model's ability to capture intricate details while maintaining computational efficiency is crucial for achieving high accuracy in license plate recognition. Additionally, the empirical evidence presented in the study, comparing various YOLO-based versions, underscores YOLOv6's superiority in terms of accuracy and computational efficiency.

Moreover, YOLOv6 addresses the critical balance between accuracy and speed, a key requirement for real-time applications such as license plate recognition. The enhancements in the model's architecture and algorithmic improvements contribute to faster processing times without compromising accuracy, making it well-suited for applications that demand both high precision and low computation costs.

As results, YOLOv6's achievement of better results in terms of high accuracy and low computation cost can be attributed to its innovative architecture, grid-based analysis, advancements in deep learning algorithms, and a meticulous design that prioritizes the critical balance between accuracy and speed. The empirical comparisons presented in the study validate these claims, providing tangible evidence of YOLOv6's effectiveness in meeting the requirements of accurate and efficient license plate recognition.

V. Conclusion

This research paper highlights the significance of License Plate Recognition (LPR) in IoT smart parking systems. It explores the use of traditional and deep learning-based methods for LPR detection, with a focus on why deep learning approaches are preferred. The challenges faced by deep learning-based LPR methods, particularly accuracy rate and computation cost, are addressed using insights from previous studies. The proposed method introduces a solution based on the YOLOv6 algorithm to overcome these challenges. The process entails creating a custom dataset, carrying out testing, validation, and training, and assessing the suggested model's performance. The experimental results demonstrate that the suggested method outperforms existing state-of-the-art methods, offering improved accuracy and reduced computation costs. The research contributes to enhancing LPR technology in IoT smart parking systems, enabling more efficient and reliable parking management solutions. For future work, an interesting direction would be to extend the proposed YOLOv6-based LPR system to incorporate real-time tracking of vehicles within the smart parking system. This would enable continuous monitoring and tracking of vehicles, providing valuable information for parking space availability and management. Cross-Domain License Plate Recognition: Another potential avenue for future research is to explore the applicability of the proposed method in cross-domain scenarios. Investigating the adaptation of the YOLOv6-based LPR system to different environments, such as night-time conditions, adverse weather, or different camera viewpoints, would contribute to the robustness and generalization capabilities of the model. This would expand the practical use cases of the system beyond traditional smart parking systems.

References

[1] F. Al-Turjman and A. Malekloo, "Smart parking in IoT-enabled cities: A survey," Sustainable Cities and Society, vol. 49, p. 101608, 2019.

[2] L. F. Luque-Vega, D. A. Michel-Torres, E. Lopez-Neri, M. A. Carlos-Mancilla, and L. E. González-Jiménez, "Iot smart parking system based on the visual-aided smart vehicle presence sensor: SPIN-V," Sensors, vol. 20, no. 5, p. 1476, 2020.

[3] K. S. Awaisi et al., "Towards a fog enabled efficient car parking architecture," IEEE Access, vol. 7, pp. 159100-159111, 2019.

[4] Y. Agarwal, P. Ratnani, U. Shah, and P. Jain, "IoT based smart parking system," in 2021 5th international conference on intelligent computing and control systems (ICICCS), 2021: IEEE, pp. 464-470.

[5] S. B. Atitallah, M. Driss, W. Boulila, and H. B. Ghézala, "Leveraging Deep Learning and IoT big data analytics to support the smart cities development: Review and future directions," Computer Science Review, vol. 38, p. 100303, 2020.

[6] M. G. Diaz Ogás, R. Fabregat, and S. Aciar, "Survey of smart parking systems," Applied Sciences, vol. 10, no. 11, p. 3872, 2020.

[7] A. A. Mei Choo Ang, Kok Weng Ng, Elankovan Sundararajan, Marzieh Mogharrebi, Teck Loon Lim, "Multi-core Frameworks Investigation on A Real-Time Object Tracking Application," Journal of Theoretical & Applied Information Technology, 2014.

[8] H. Mohapatra and A. K. Rath, "An IoT based efficient multi-objective real-time smart parking system," International journal of sensor networks, vol. 37, no. 4, pp. 219-232, 2021.

[9] W. Weihong and T. Jiaoyang, "Research on license plate recognition algorithms based on deep learning in complex environment," IEEE Access, vol. 8, pp. 91661-91675, 2020.

[10] Lubna, N. Mufti, and S. A. A. Shah, "Automatic number plate Recognition: A detailed survey of relevant algorithms," Sensors, vol. 21, no. 9, p. 3028, 2021.

[11] G. Ali et al., "IoT based smart parking system using deep long short memory network," Electronics, vol. 9, no. 10, p. 1696, 2020.

[12] A. Aghamohammadi et al., "A deep learning model for ergonomics risk assessment and sports and health monitoring in self-occluded images," Signal, Image and Video Processing, pp. 1-13, 2023.

[13] X. Wu, D. Sahoo, and S. C. Hoi, "Recent advances in deep learning for object detection," Neurocomputing, vol. 396, pp. 39-64, 2020.

[14] P. Kaur, Y. Kumar, S. Ahmed, A. Alhumam, R. Singla, and M. F. Ijaz, "Automatic License Plate Recognition System for Vehicles Using a CNN," Computers, Materials & Continua, vol. 71, no. 1, 2022.

[15] D. N. C. Loong, S. Isaak, and Y. Yusof, "Machine vision based smart parking system using Internet of Things," Telkomnika (Telecommunication Computing Electronics and Control), vol. 17, no. 4, pp. 2098-2106, 2019.

[16] G. Manjula, G. Govinda Rajulu, R. Anand, and J. Thirukrishna, "Implementation of smart parking application using IoT and machine learning algorithms," in Computer Networks and Inventive Communication Technologies: Proceedings of Fourth ICCNCT 2021, 2022: Springer, pp. 247-257.

[17] A. T. Kabir et al., "An IoT based intelligent parking system for the unutilized parking area with real-time monitoring using mobile and web application," in 2021 International Conference on Intelligent Technologies (CONIT), 2021: IEEE, pp. 1-7.

[18] P. Amarasooriya and M. Peiris, "Implementation of Smart Parking System Using Image Processing," MPPL, Implementation of Smart Parking System Using Image Processing (June 10, 2023), 2023.

[19] C. Silpa, "Implementing Deep Learning algorithms for Automatic license plate Recognition," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 14, no. 2, pp. 203-212, 2023.

[20] C. Gupta, N. S. Gill, P. Gulia, and J. M. Chatterjee, "A novel finetuned YOLOv6 transfer learning model for real-time object detection," Journal of Real-Time Image Processing, vol. 20, no. 3, p. 42, 2023.

[21] C. Li et al., "YOLOv6: A single-stage object detection framework for industrial applications," arXiv preprint arXiv:2209.02976, 2022.

[22] S. Rath. "YOLOv6 Object Detection – Paper Explanation and Inference." https://learnopencv.com/yolov6-object-detection/#YOLOv6-Model-Architecture (accessed.

[23] S. Norkobil Saydirasulovich, A. Abdusalomov, M. K. Jamil, R. Nasimov, D. Kozhamzharova, and Y.-I. Cho, "A YOLOv6-based improved fire detection approach for smart city environments," Sensors, vol. 23, no. 6, p. 3161, 2023.

Influence of Membership Function and Degree on Sorghum Growth Prediction Models in Machine Learning

Abdul Rahman[1], Ermatita[2*], Dedik Budianta[3], Abdiansah[4]
Doctoral Program in Engineering Science, Universitas Sriwijaya, Palembang, Indonesia[1]
Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia[2, 4]
Faculty of Agricultural, Universitas Sriwijaya, Palembang, Indonesia[3]
Faculty of Computer Science and Engineering, Universitas Multi Data, Palembang, Indonesia[1]

***Abstract*—Rapid advances in science and technology have significantly changed plant growth modeling. The main contribution to this transformation lies in using Machine Learning (ML) techniques. This study focuses on sorghum, an important agricultural crop with significant economic implications. Crop yield studies include temperature, humidity, climate, rainfall, and soil nutrition. This research has a novelty: the input factors for predicting sorghum plant growth, namely the treatment of applying organic fertilizer and dolomite lime to sorghum planting land. The three predicted sorghum plant growth factors, namely Height, Biomass, and Panicle weight, are the reasons for using the Multiple Adaptive Neural Fuzzy Inference System (MANFIS) model. This research investigates the impact of Membership Function and Degree on the MANFIS model. A comprehensive comparison of various membership functions, including Gaussian, Triangular, Bell, and Trapezoidal functions, along with various degrees of membership, has been carried out. The dataset used includes data related to sorghum growth obtained from field experiments. The main objective was to assess the effectiveness of membership and degree functions in accurately predicting sorghum growth parameters, consisting of height, biomass, and panicle weight. This assessment uses metrics such as MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), and RMSE (Root Mean Square Error) to evaluate the predictive performance of the MANFIS model when using four different types of membership functions and degrees. The results obtained the best level of accuracy in predicting panicle weight (ANFIS-3) with chicken manure treatment using the Trapezoidal membership function type and degree of membership function [3,3] with MAPE results of 5.77%, MAE of 0.2994, and RMSE of 0.395.**

***Keywords*—Prediction; MANFIS; membership function; organic fertilizer; sorghum**

I. Introduction

Rice is the staple food of the Indonesian population; rice production in Indonesia in the last five years (2018-2022) has continued to decline, from 59.2 million tons in 2018 to 54.74 million tons in 2022, as well as the harvested area which in 2018 reached 11.37 million ha to 10.45 million ha in 2022[1]. Therefore, alternative food is needed to replace rice to ensure food security in Indonesia. *Sorghum* is an alternative crop suitable for planting in less fertile areas, such as tidal land. Suboptimal tidal swamp land has low fertility, an acidic pH, and low nutrients [2]. Sorghum plants can also be used for food diversification other than rice to maintain food security in Indonesia. Sorghum is more drought tolerant than similar crops such as corn and wheat [3]. Sorghum is suitable for cultivation in Indonesia because of its drought tolerance and adaptability to tropical areas [4].

The main problems of plant growth in tidal land are the level of water saturation and anaerobic conditions in the rhizosphere, pyrite or sulfide materials found in the soil, toxicity of Al, Fe, and Mn; highly acidic soil reaction, and low content of N, P, K, Ca, and Mg [5] [6] [7] [8]. Enhancing soil fertility in tidal land areas can be achieved through the application of fertilizers [9]. In such regions, leveraging local resources for sustainable agricultural practices is essential. One viable option is the utilization of organic fertilizers, such as chicken manure, cow manure, and vermicompost [10]. These locally available resources provide essential nutrients to the soil and promote soil health and microbial activity. By adopting a strategy that combines the application of these organic fertilizers with tidal land management practices, farmers can effectively improve soil fertility while minimizing the environmental impact, contributing to both agricultural productivity and environmental sustainability [11].

Many further research studies within the sorghum domain utilize machine learning techniques. These encompass efforts to predict sorghum biomass [12], detect and measure sorghum head counts [13], and make estimations of sorghum crop yields through machine learning algorithms [14] [15]. In combination, these investigations underscore the versatility and potential of machine learning in advancing various aspects of sorghum farming and administration, promoting more effective and sustainable agricultural practices.

Recent research on crop yield predictions utilizing machine learning techniques has emphasized the incorporation of several key input parameters. Rainfall has been identified as a significant factor in crop yield prediction, with multiple studies exploring its impact [16] [17] [18] [19], Temperature, humidity, and climate have also emerged as primary concerns, with some studies employing multiple linear regression to analyze weather forecasts by considering these parameters [16] [20] [18] [19] [21]. Moreover, soil pH and irrigation, as determinants of soil quality and optimal irrigation, are integral components of crop yield prediction models [19] [22] [21]. Wind speed, an external factor affecting plant growth, is considered in several research endeavors [20] [21]. Additionally, the location of crops is recognized as a crucial parameter in crop yield prediction, with crop location data serving as a variable in predictive analyses [18] [23]. These recent studies combine machine learning technology with a deep understanding of these diverse factors to enhance the accuracy of crop yield predictions. In this research, we introduce novelty by utilizing input parameters that include doses of organic fertilizers (chicken manure, cow manure, and vermicompost) and dolomite. Interestingly, these parameters have not been explored in previous studies related to predicting plant growth using machine learning technology.

Research conducted on predicting sorghum crop yields involve the use of various techniques, including the development of machine learning-based models. Several approaches have been employed, such as using TensorFlow with Convolutional Neural Networks (CNN) and Linear Regression to detect and estimate the weight of sorghum heads in images [24]. Additionally, a neuro-fuzzy model has been developed to predict the production rate of colorant extract from sorghum bicolor [25]. There has also been the development of image segmentation algorithms using deep learning CNN to detect and count sorghum heads[26]. Furthermore, frameworks and models have been created to detect leafhopper infestations in sorghum plants using deep learning technology and the YOLOv5m model [27]. Moreover, a performance evaluation of three deep learning methods, namely EfficientDet, SSD, and YOLOv4, has been conducted for the detection of sorghum heads in UAV RGB image [28].

In this research, a machine learning prediction model was developed using the Multiple Adaptive Fuzzy Inference System (MANFIS) model. ANFIS harnesses the strengths of neural networks and fuzzy logic, collectively enhancing its predictive capabilities [29]. This fusion of Neural Network and Fuzzy Logic equips ANFIS with a robust framework for precise predictions. Additionally, ANFIS possesses the unique capability to integrate both numerical and linguistic knowledge, making it adaptable and valuable across a broad spectrum of domains [30]. Its capacity to handle diverse types of information is a substantial asset, enabling it to efficiently address a multitude of real-world situations. However, integrating ANFIS models into soil remediation offers a promising avenue for restoring contaminated land to a fertile state, enabling sustainable agricultural practices. Many have widely implemented the ANFIS for prediction, classification, and clustering [31] [32] [33]. In this study, the input parameters and output parameters have more than one parameter. Hence, the prediction model in this study uses nine ANFIS models to predict three output parameters (height, biomass, and panicle weight) of sorghum plants with three different organic fertilizer treatment datasets.

The connection between membership functions and the degree of membership in ANFIS is essential for fuzzy logic-based inference. Here, the functions establish fuzzy sets, and the degree of membership measures how closely input values align with these sets. This association is vital for the fuzzy reasoning and decision-making procedures in the ANFIS framework. In the selection of membership functions and degrees of membership in the ANFIS model, research has highlighted the importance of choosing appropriate membership functions. Studies have explored various forms of membership functions, such as triangular, trapezoidal, shape-bell, and Gaussian, as well as the selection of the correct number of membership functions [34]. The choice of appropriate membership functions can have a significant impact on the performance of the ANFIS model [35]. Experiments are often employed to determine the optimal membership functions, allowing precise adjustments to the system's performance [36]. Over time, research continues to develop best practices in the selection of membership functions and degrees of membership to enhance the performance of ANFIS-based systems. Therefore, in this study, to optimize the accuracy of prediction results using the MANFIS model, the selection of membership function types and degrees of membership is conducted on nine ANFIS models. Four membership functions and four combinations of degrees of membership will be evaluated for each input and output parameter of the MANFIS model designed to obtain the best prediction results for sorghum plant growth.

This research advances the field of plant growth prediction through machine learning by introducing new input factors, such as organic fertilizer dosage and dolomite, which previous studies have not explored. In this study, we present novelty by including input variables such as the amount of organic fertilizer (chicken manure, cow manure, and vermicompost) and dolomite. Previous studies on predicting plant growth using machine learning technology did not investigate these variables. Additionally, it emphasizes the importance of carefully choosing appropriate membership functions for the ANFIS model and conducting experiments to improve its performance. These efforts contribute to developing best practices in this domain, enhancing the accuracy and effectiveness of predicting crop yields.

II. Related Works

The choice of membership function significantly influences the prediction model's accuracy when using ANFIS. One study compared eight different ANFIS membership functions to optimize ERP satisfaction values, ultimately revealing that the triangular membership function yielded the best prediction results [37]. To guarantee reliable and accurate predictions, the research prioritized two crucial factors: the number of inputs within the training dataset and the selection of the membership function within the ANFIS model. This optimization procedure included comparing outcomes with techniques like Particle Swarm Optimization and Genetic Algorithms [36]. Furthermore, researchers conducted performance evaluations to assess how effectively the ANFIS model addressed various classification problems by investigating four popular forms of membership functions: triangular, bell-shaped, trapezoidal, and Gaussian [35]. Another study delved into the impact of different membership function types, specifically triangular, trapezoidal, and Gaussian, on the performance of a fuzzy logic controller [38]. In a different context, researchers employed two approaches to generate Gaussian and triangular fuzzy membership functions using fuzzy c-means for predicting sunspots [39]. These various investigations collectively contribute to our understanding of the significance of membership functions and their impact on ANFIS model performance in different applications. In this research, a comparison was carried out among four types of membership functions and four combinations of membership degree functions across nine ANFIS models employed in a machine learning framework for predicting the growth of sorghum. This comparison aims to determine the most accurate prediction results within the constructed model.

The ANFIS model utilizes the Takagi-Sugeno rule set for its fuzzy inference system. Eq. (1) and Eq. (2) present a standard rule set for the commonly used first-order Takagi-Sugeno fuzzy model, which includes two fuzzy if-then rules [40].

Rule 1: If (x is A_1) and (y is B_1), then: $Z_1 = p_1x + q_1y + r_1$ (1)

Rule 2: If (x is A_2) and (y is B_2), then: $Z_2 = p_2x + q_2y+ r_2$ (2)

where, p_1, p_2, q_1, q_2, r_1, and r_2 are linear, and A_1, A_2, B_1, and B_2 are non-linear parameters.

In Fig. 1 shows the structure of ANFIS, which consists of five layers [41]. The framework of the ANFIS method has 5 (five) layers, namely the fuzzification layer, the rule layer, the normalization layer, the defuzzification layer, and a single neuro result [42] [43].

- Layer 1: This layer serves as the fuzzification layer, where each neuron's output corresponds to the degree of membership determined by the input membership function. The fundamental categories of membership functions include four types: triangular, trapezoidal, bell-shaped, and Gaussian [35].

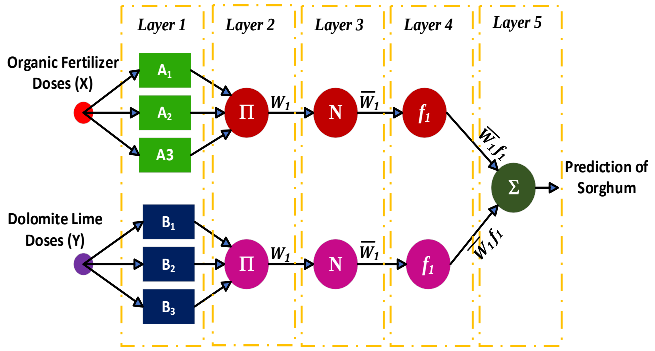


Fig. 1. A schematic of an ANFIS structure.

The triangular membership function (Trimf) stands out as the most straightforward among the various membership functions. It requires three parameters to define the three points, as illustrated in Eq. (3).

$$Trimf(x,a,b,c) = max\left(min\left(\frac{x-a}{b-a},\frac{c-x}{c-b},0\right)\right) \quad (3)$$

where the value a < b < c which represents the coordinates of the Trimf on the x-axis.

Eq. (4) illustrates how four scalar parameters define the curve of the trapezoidal membership function (Trapmf).

$$Trapmf(x,a,b,c,d) = max\left(min\left(\frac{x-a}{b-a},1,\frac{d-x}{d-b},0\right)\right) \quad (4)$$

where the value a < b < c < d which represents the coordinates of the Trapmf on the x-axis

The general bell-shaped membership function (Gbellmf) features a symmetric shape resembling a bell, as illustrated in Eq. (5).

$$Gbellmf(x,a,b,c) = \frac{1}{1+\left[\frac{x-c}{a}\right]^{2b}} \quad (5)$$

where, c is the center of the curve in the universe of speech, a determines the width of the bell-shaped curve, and b is a positive integer.

The Gaussian membership function (Gaussmf) relies on two parameters: c to locate the center and σ to specify the curve's width, as shown in Eq. (6).

$$Gaussmf(c,\sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \quad (6)$$

where, c is the center of the cluster and width value σ are used to describe the Gaussmf.

- Layer 2: This layer comprises a constant neuron (represented by the symbol Π), which computes the product of all input values, as indicated in Eq. (7).

$$\boldsymbol{wk} = \boldsymbol{\mu Ak} \,.\, \boldsymbol{\mu Bk} \quad (7)$$

Typically, practitioners use the AND operator and refer to the result of this computation as the firing strength of a rule. Each neuron corresponds to a specific rule indexed ask.

- Layer 3: Each neuron in this layer is a constant neuron, represented by the *N*. The calculation takes the *k* firing strength (*wk*) ratio to the total sum of firing strengths in the second layer, as shown in Eq. (8).

$$\overline{w}_k = w_k w_1 + w_2, \quad i = 1,2 \quad (8)$$

The result obtained from this calculation is termed the normalized firing strength.

- Layer 4: This layer consists of neurons that adapt to an output, as shown in Eq. (9).

$$\overline{w}_k f_k = \overline{w}_k \langle q_k Z_{t-1} + r_k Z_{t-2} + s_k \rangle \quad (9)$$

where, $\overline{w}_k$ is the normalized firing strength in the third layer and qk, rk, and sk are the parameters of the neuron. These parameters are commonly called consequent parameters.

- Layer 5: This layer comprises a solitary neuron (represented by a symbol) that results from summating all outputs from the fourth layer, as depicted in Eq. (10).

$$\sum \overline{w}_k f_k = \frac{\sum_k w_k f_k}{\sum_k w_k} \quad (10)$$

A. *Proposed Method*

The research methodology aims to predict Sorghum growth using MANFIS (Multiple Adaptive Neuro-Fuzzy Inference System) models, emphasizing optimizing the selection of membership functions and membership degrees to achieve the highest accuracy, as depicted in Fig. 2. Data collection involved conducting experiments that utilized three types of organic fertilizers: chicken manure, cow manure, and vermicompost, in combination with dolomite lime on tidal soil. The experimental design employed a two-factor factorial design. The data obtained from these experiments served as the MANFIS model's dataset to predict three sorghum growth parameters: height, biomass weight, and panicle weight. The dosage of organic fertilizers and dolomite lime forms the basis for these predictions.

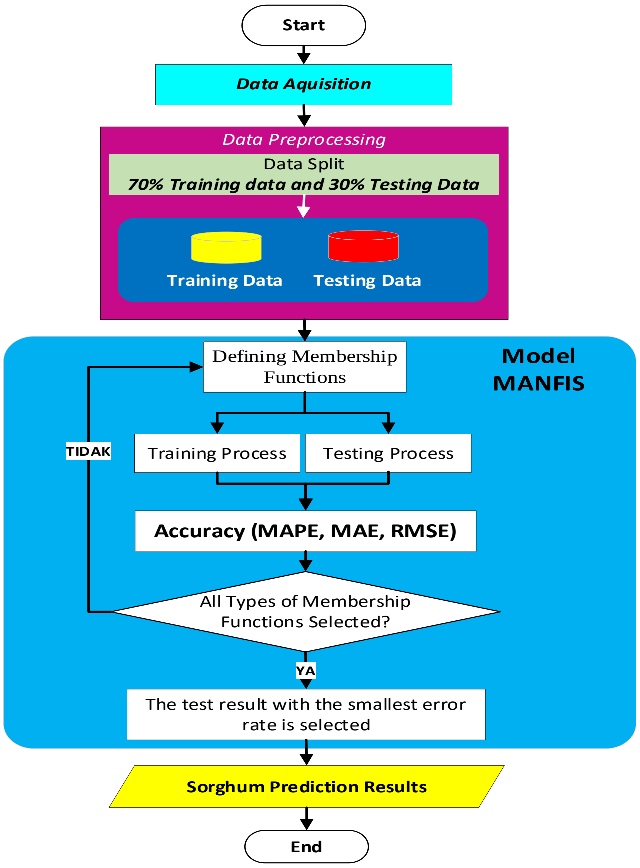

Start
Data Aquisition
Data Preprocessing
Data Split
70% Training data and 30% Testing Data
Training Data
Testing Data
Model MANFIS
Defining Membership Functions
TIDAK
Training Process
Testing Process
Accuracy (MAPE, MAE, RMSE)
All Types of Membership Functions Selected?
YA
The test result with the smallest error rate is selected
Sorghum Prediction Results
End

Fig. 2. Research methodology.

The initial phase of our research began with problem identification, focusing on utilizing organic fertilizer comprising chicken manure, cow manure, and vermicompost, which was applied to tidal lands to nurture sorghum plants. Subsequently, we collected data, drawing information from the growth records of sorghum plants cultivated on tidal lands subjected to three different organic fertilizer treatments or doses.

1) Vermicompost fertilizer: The treatment parameter for vermicompost fertilizer used four doses, namely: 0, 2.5, 5, and 7.5 tons/ha, combined with two doses of dolomite lime (0 and 0.404 tons/ha) and repeated three times.

2) Chicken manure: The treatment parameters of chicken manure used four doses, namely: 0, 5, 6.5, and 8.5 tons/ha, combined with two doses of dolomite lime (0 and 0.404 tons/ha), and repeated three times.

3) Cow manure: The treatment parameter for cow manure used four doses, namely: 0, 5, 10, and 15 tons/ha, combined with two doses of dolomite lime (0 and 1.84 tons/ha), and repeated three times.

The researchers obtained the dataset by conducting experimental sorghum cultivation in a greenhouse. We applied organic fertilizer to the cultivated land two weeks before planting sorghum. After 105 days, we harvested the sorghum plants and measured three growth parameters: height, biomass, and panicle weight.

In the subsequent phase, we conducted data training using four different membership functions, namely triangular, trapezoidal, bell-shaped, and gaussian, to achieve the highest level of accuracy. We utilized 70% of the entire dataset in the training phase and executed the training process on nine ANFIS models. Fig. 3 illustrates the configuration of the MANFIS model designed for chicken manure fertilizer treatment. This model comprises three ANFIS components, denoted ANFIS-1, ANFIS-2, and ANFIS-3. The input variables for their membership functions are characterized by degrees of membership, specifically (4, 2). In particular, the dosage of chicken manure fertilizer has four membership degrees, while dolomite lime has two membership degrees.

In this process, the training data, randomly selected from a dataset, is employed to train individual ANFIS models using distinct types and degrees of membership functions. There are four distinct types of membership functions and four combinations of membership function degrees applied during this training phase. The ANFIS model training utilizes these combinations of membership function types and degrees, including Triangular, Trapezoidal, Generalized bell-shaped, and Gaussian functions, along with four degrees of membership functions: {3,2}, {3,3}, {2,2}, and {4,2}.

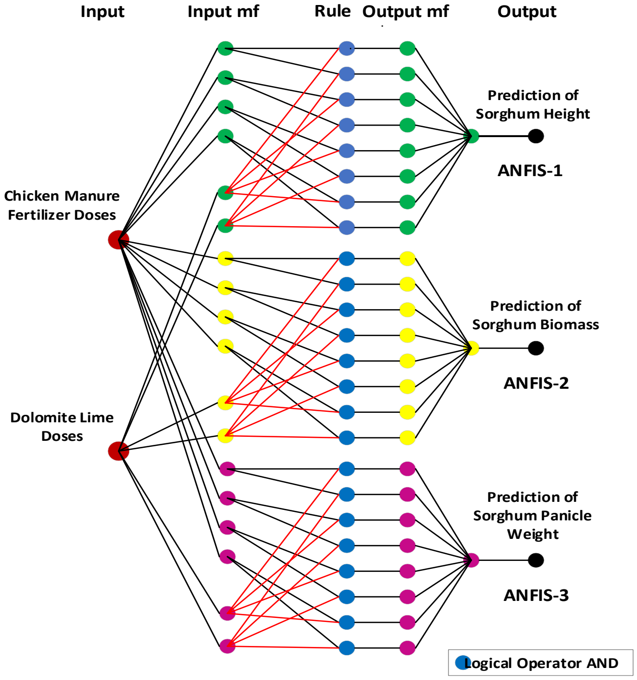

Input
Input mf
Rule
Output mf
Output
Chicken Manure Fertilizer Doses
Dolomite Lime Doses
Prediction of Sorghum Height
ANFIS-1
Prediction of Sorghum Biomass
ANFIS-2
Prediction of Sorghum Panicle Weight
ANFIS-3
Logical Operator AND

Fig. 3. Model structure for chicken manure treatment.

Using these diverse combinations of membership function types and degrees, we assess the accuracy of the prediction outcomes for each ANFIS model using three accuracy measurement indicators: Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). We derive the formulas for calculating the accuracy of prediction results from Eq. (11), (12), and (13) as detailed in reference [44].

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{A_t - P_t}{A_t}\right] x\ 100\% \quad (11)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}[P_t - A_t] \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_t - P_t)^2} \quad (13)$$

where, A_t represents the target value, P_t refers to the prediction value (the model's output), and n is the number of data points.

III. Result and Analysis

The results of the ANFIS model's accuracy evaluation, including three prediction accuracy metrics (MAPE, MAE, and RMSE), along with the utilization of four types of membership functions and four degrees of membership functions for each treatment related to the application of organic fertilizer and dolomite lime, have been recorded in Tables I, II, and III. These results specifically relate to three observed output variables: the sorghum plant's height, biomass weight, and panicle weight.

Table I presents the accuracy assessment results for MANFIS models (ANFIS-1, ANFIS-2, and ANFIS-3) using the metrics MAPE, MAE, and RMSE. These results stem from the training and testing processes conducted on data from the chicken manure treatment dataset. To streamline the process of identifying the membership function types and degrees that produce the most accurate results, we visually represent the accuracy measurements in Table I through Fig. 4, Fig. 5, and Fig. 6.

Fig. 4 visually depicts the accuracy outcomes in the context of chicken manure treatment, explicitly focusing on sorghum height (ANFIS-1) as the output parameter. It is evident from Fig. 4 that the highest accuracy values, as measured by the three accuracy assessment tools, are achieved when using the trapezoidal membership function type and membership function degrees {2, 2}. The corresponding accuracy values are MAPE of 13.23%, MAE of 11.1969, and RMSE of 14.7685.

Fig. 5 provides a visual representation of the accuracy results in the context of chicken manure treatment, explicitly highlighting the Biomass weight (ANFIS-2) of sorghum as the output parameter. The figure demonstrates that the highest accuracy values, as assessed by the three accuracy measurement tools, are attained when employing the Bell-shaped membership function type with membership function degrees {4, 2}. The corresponding accuracy values are MAPE of 20.89%, MAE of 2.5163, and RMSE of 3.2552.

TABLE I. Accuracy Measurement Results of ANFIS for Chicken Manure Treatment

Membership Function		**Output Parameters**								
Type	**Degree**	**Sorghum Height (ANFIS-1)**			**Sorghum Biomass Weight (ANFIS-2)**			**Sorghum Panicle Weight (ANFIS-3)**		
		MAPE	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**
Triangular	[3 3]	7.57	7.6624	9.6971	24.77	3.8184	4.7042	7.07	0.3216	0.4717
	[3 2]	7.11	7.0251	9.5958	27.68	4.0041	4.8796	6.59	0.3109	0.4947
	[2 2]	8.05	8.0693	10.3974	26.63	3.7491	4.9433	6.31	0.2929	0.4766
	[4 2]	7.44	7.33	9.9284	21.68	3.6174	5.1553	6.8	0.3148	0.4809
Trapezoidal	[3 3]	7.34	7.3734	9.4516	25.43	3.64	4.8043	7.25	0.3321	0.4718
	[3 2]	7.64	7.6231	9.6295	25.12	3.7713	4.6919	6.61	0.3146	0.4957
	[2 2]	7.29	7.2511	9.5181	21.62	3.6038	5.0974	6.49	0.3076	0.4847
	[4 2]	7.32	7.3311	9.5143	22	3.6717	5.2302	7.45	0.3784	0.6374
Bell-shaped	[3 3]	8.25	8.9019	13.2622	30.17	4.2531	5.4302	15.93	0.6813	0.8268
	[3 2]	7.38	7.1854	10.9684	26.87	3.9804	5.0931	13.7	0.5578	0.6737
	[2 2]	8.16	8.2203	10.2186	28.69	4.1439	5.2805	13.42	0.5579	0.6636
	[4 2]	7.14	6.9103	9.9857	27.93	4.2529	5.4921	14.06	0.5899	0.6771
Gaussian	[3 3]	7.22	7.0862	9.8436	35.13	5.34	6.9752	6.34	0.2997	0.4798
	[3 2]	7.55	7.5134	9.5617	25.51	3.7674	4.785	6.66	0.3145	0.4511
	[2 2]	7.8	7.8058	9.933	35.28	5.5481	7.1964	6.55	0.3087	0.4533
	[4 2]	7.37	7.3733	10.1953	27.44	4.0184	4.8416	14.11	0.5176	0.7754

TABLE II. ACCURACY MEASUREMENT RESULTS OF ANFIS FOR COW MANURE TREATMENT

Membership Function		Output Parameters								
Type	Degree	Sorghum Height (ANFIS-4)			Sorghum Biomass Weight (ANFIS-5)			Sorghum Panicle Weight (ANFIS-6)		
		MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE
Triangular	[3 3]	7.57	7.6624	9.6971	24.77	3.8184	4.7042	7.07	0.3216	0.4717
	[3 2]	7.11	7.0251	9.5958	27.68	4.0041	4.8796	6.59	0.3109	0.4947
	[2 2]	8.05	8.0693	10.3974	26.63	3.7491	4.9433	6.31	0.2929	0.4766
	[4 2]	7.44	7.33	9.9284	21.68	3.6174	5.1553	6.8	0.3148	0.4809
Trapezoidal	[3 3]	7.34	7.3734	9.4516	25.43	3.64	4.8043	7.25	0.3321	0.4718
	[3 2]	7.64	7.6231	9.6295	25.12	3.7713	4.6919	6.61	0.3146	0.4957
	[2 2]	7.29	7.2511	9.5181	21.62	3.6038	5.0974	6.49	0.3076	0.4847
	[4 2]	7.32	7.3311	9.5143	22	3.6717	5.2302	7.45	0.3784	0.6374
Bell-shaped	[3 3]	8.25	8.9019	13.2622	30.17	4.2531	5.4302	15.93	0.6813	0.8268
	[3 2]	7.38	7.1854	10.9684	26.87	3.9804	5.0931	13.7	0.5578	0.6737
	[2 2]	8.16	8.2203	10.2186	28.69	4.1439	5.2805	13.42	0.5579	0.6636
	[4 2]	7.14	6.9103	9.9857	27.93	4.2529	5.4921	14.06	0.5899	0.6771
Gaussian	[3 3]	7.22	7.0862	9.8436	35.13	5.34	6.9752	6.34	0.2997	0.4798
	[3 2]	7.55	7.5134	9.5617	25.51	3.7674	4.785	6.66	0.3145	0.4511
	[2 2]	7.8	7.8058	9.933	35.28	5.5481	7.1964	6.55	0.3087	0.4533
	[4 2]	7.37	7.3733	10.1953	27.44	4.0184	4.8416	14.11	0.5176	0.7754

TABLE III. ACCURACY MEASUREMENT RESULTS OF ANFIS FOR VERMICOMPOST TREATMENT

Membership Function		Output Parameters								
Type	Degree	Sorghum Height (ANFIS-7)			Sorghum Biomass Weight (ANFIS-8)			Sorghum Panicle Weight (ANFIS-9)		
		MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE
Triangular	[3 3]	6.05	5.8729	8.2177	21.72	4.9165	10.9312	7.49	0.4447	0.5465
	[3 2]	6.43	6.2208	8.1537	34.84	5.7267	9.8537	7.7	0.4538	0.5528
	[2 2]	6.48	6.4189	8.6914	23.26	4.8663	10.7253	8.09	0.4715	0.5574
	[4 2]	6.49	6.5281	8.6078	30.95	5.4974	9.5147	8.9	0.5109	0.6914
Trapezoidal	[3 3]	6.72	6.372	9.0279	25.45	5.1818	11.3331	7.46	0.4481	0.5494
	[3 2]	6.26	6.0985	7.8162	34.35	5.7109	9.8386	8.08	0.4779	0.5521
	[2 2]	6.35	6.3306	8.5164	34.48	5.7302	9.8525	7.75	0.4592	0.617
	[4 2]	6.15	5.9474	7.9856	50.26	6.9171	15.6322	7.25	0.4357	0.6138
Bell-shaped	[3 3]	6.34	6.3658	9.3784	36.13	6.5497	10.5085	7.79	0.4646	0.6032
	[3 2]	6.42	6.2479	7.9644	31.06	5.936	10.1495	7.74	0.4584	0.5958
	[2 2]	6.97	6.6223	8.85	23.91	4.9333	10.683	7.5	0.4429	0.6034
	[4 2]	6.44	6.2097	8.1773	22.62	4.8991	10.8324	7.97	0.47	0.6102
Gaussian	[3 3]	6.28	6.0763	7.8944	22.94	4.9947	11.3718	7.98	0.4676	0.5786
	[3 2]	6.18	5.9994	7.8711	32.32	5.6328	9.6765	7.73	0.4588	0.5605
	[2 2]	6.08	5.9147	7.7382	22.06	4.9958	11.3755	7.89	0.4641	0.5479
	[4 2]	6.43	6.2469	8.039	32.95	5.6962	9.6584	7.5	0.4522	0.5749

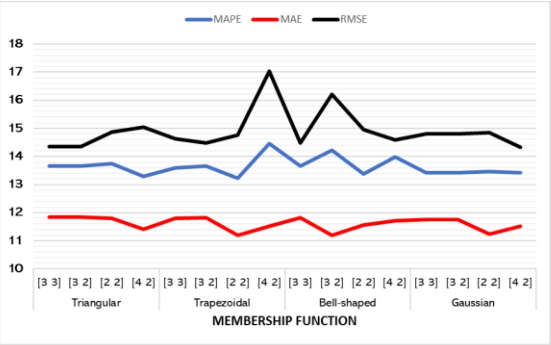


Fig. 4. The prediction accuracy level of sorghum height in chicken manure treatment.

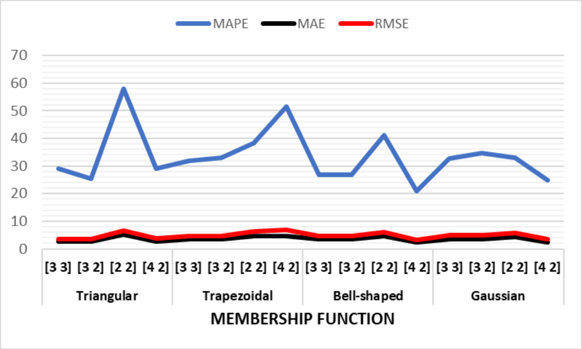


Fig. 5. The prediction accuracy level of sorghum biomass in chicken manure treatment.

Fig. 6 visually represents the accuracy outcomes in the context of chicken manure treatment, explicitly focusing on sorghum's Panicle weight (ANFIS-3) as the output parameter. Fig. 6 shows that the highest accuracy values, as measured by the three accuracy assessment tools, are achieved when using the Trapezoidal membership function type and membership function degrees {3, 3}. The corresponding accuracy values are MAPE of 5.77%, MAE of 0.2994, and RMSE of 0.395.

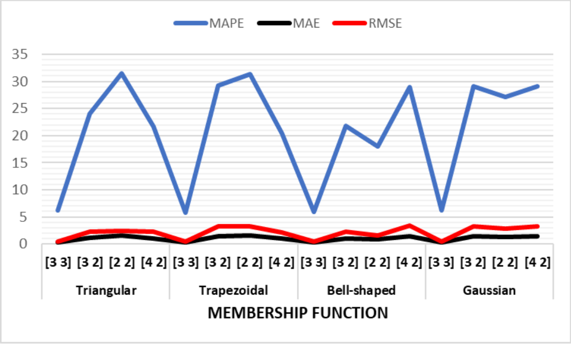


Fig. 6. The prediction accuracy level of sorghum panicle weight in chicken manure treatment.

To identify the membership function types and degrees that offer the highest accuracy in treatments involving cow manure and vermicompost, a similar methodology was applied to that used in the chicken manure treatment. By assessing the accuracy using MAPE, MAE, and RMSE for three distinct organic fertilizer dosages in a tidal swamp area for sorghum growth, the specific membership function types, and degrees for the MANFIS model predicting sorghum growth are determined and outlined in Table IV.

TABLE IV. RESULTS OF MEMBERSHIP FUNCTION TYPE AND DEGREE SELECTION

MANFIS	Type Membership Function	Degree Membership Function
ANFIS-1	Trapezoidal	{2,2}
ANFIS-2	Bell-Shaped	{4,2}
ANFIS-3	Trapezoidal	{3,3}
ANFIS-4	Triangular	{3,2}
ANFIS-5	Trapezoidal	{2,2}
ANFIS-6	Triangular	{2,2}
ANFIS-7	Triangular	{3,3}
ANFIS-8	Triangular	{3,3}
ANFIS-9	Trapezoidal	{4,2}

Based on the accuracy testing results achieved through the utilization of four types of membership functions and four degrees of membership functions in the MANFIS model, as outlined in Table IV, we depict the schematic of the MANFIS model designed for predicting three sorghum plant growth parameters in tidal swamp land in Fig. 7.

Subsequently, the predefined MANFIS model, with the chosen membership function type and degree, is subjected to simulation using the Simulink tool. This simulation aims to predict sorghum plant growth based on input parameters related to organic fertilizer dosage and dolomite lime application. The MANFIS model is simulated in this study using the Matlab/Simulink tool, as shown in Fig. 8. The ANFIS simulation model loads the fuzzy inference system (fis) files, which are the result of the data training process on the ANFIS model. These files are the result of the data training process of the ANFIS model, according to the input parameters of organic fertilizer and the predicted output parameters. During this simulation, nine ANFIS models predict three output parameters, with three organic fertilizer treatments as inputs.

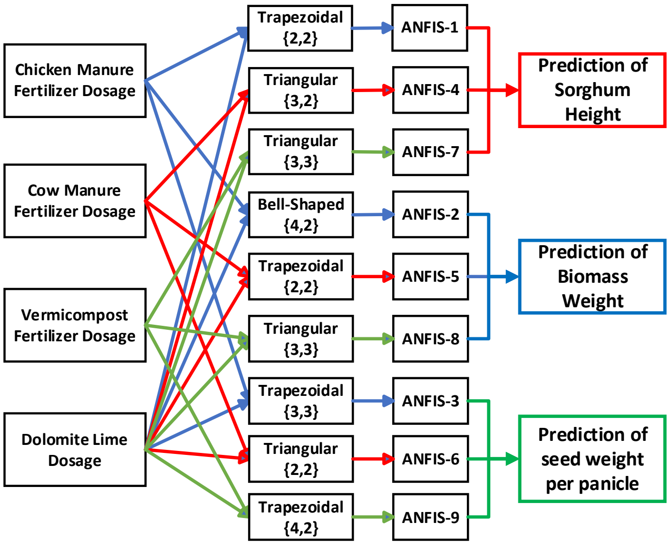


Fig. 7. MANFIS prediction model with selected membership function.

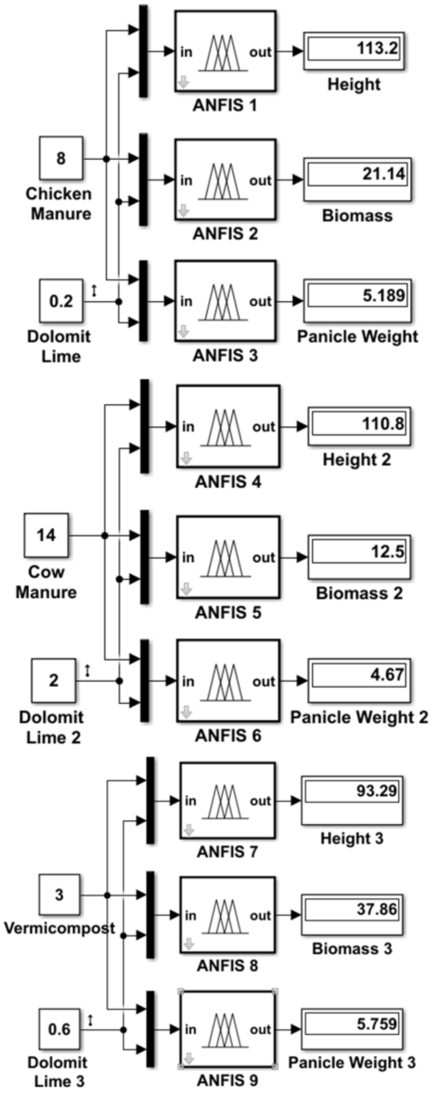


Fig. 8. Simulation of MANFIS prediction model.

In Fig. 8, for chicken manure fertilizer with a dose of 8 tons/ha and dolomite lime with a dose of 0.2 tons/ha, the predicted results are as follows: sorghum height (ANFIS-1) = 113.2 cm, sorghum biomass weight (ANFIS-2) = 21.14 tons/ha, and sorghum panicle weight (ANFIS-3) = 5.189 tons/ha. For cow manure fertilizer with a dose of 14 tons/ha and dolomite lime with a dose of 2 tons/ha, the predicted results are sorghum height (ANFIS-4) = 110.8 cm, sorghum biomass weight (ANFIS-5) = 12.5 tons/ha, and sorghum panicle weight (ANFIS-6) = 4.67 tons/ha. Similarly, for vermicompost fertilizer with a dose of 3 tons/ha and dolomite lime with a dose of 0.6 tons/ha, the predicted results are sorghum height (ANFIS-7) = 93.29 cm, sorghum biomass weight (ANFIS-8) = 37.86 tons/ha, and Sorghum panicle weight (ANFIS-9) = 5.759 tons/ha.

Table V presents the simulation results of predictions (height, biomass, and sorghum panicle weight) for various treatments with different doses of organic fertilizer and dolomite lime. In this Table V, each organic fertilizer is subjected to three combinations of organic fertilizer and dolomite lime doses, resulting in three predicted outcome parameters through the conducted simulations. This MANFIS model simulation (see Fig. 8) provides insights into the predicted outcomes obtained from applying various doses of organic fertilizer and dolomite lime on tidal land soil for sorghum plant growth prediction.

TABLE V. RESULTS OF MANFIS MODEL SIMULATION PREDICTION

Fertilizer Dosage (ton/ha)	Dolomite Lime Dosage (ton/ha)	Prediction Results		
		Height (cm)	Biomass (ton/ha)	Panicle weight (ton/ha)
Chicken Manure				
5	0.404	109.6	22.76	6.3
2	0.6	98.34	19.58	6.331
8	0.2	113.2	21.14	5.189
Cow Manure				
5	1.5	104.5	21	3.782
8	1	108.2	18.86	4.493
14	2	110.8	12.5	4.67
Vermicompost				
5	0.404	101.1	17.4	5.595
8	0.1	50.23	10.11	3.1
3	0.6	93.29	37.86	5.759

IV. CONCLUSION

The type of membership function used in this prediction model has a different type of membership function for each treatment of organic fertilizer on tidal soil for sorghum. The MANFIS model, using a dataset derived from observational data on sorghum plant height, biomass, and panicle weight with treatments of organic fertilizers (chicken manure, cow manure, vermicompost) in tidal soil, can be implemented to predict sorghum plant growth, including three predicted parameters: plant height, biomass, and panicle weight. The structure of the MANFIS model designed in this study consists of nine ANFIS models. The comparison of prediction accuracy results, utilizing three measurement tools - MAPE, MAE, and RMSE, demonstrates that the choice of membership function types and degrees influences the accuracy of prediction outcomes for each input data originating from the organic fertilizer treatment in tidal swamp land. This research achieved the highest prediction accuracy with the ANFIS-5 model for predicting the panicle weight of sorghum using Trapezoidal membership type and membership function parameters [3,3]. The model assessed the accuracy levels with a Mean Absolute Percentage Error (MAPE) of 5.77%, Mean Absolute Error (MAE) of 0.2994, and Root Mean Square Error (RMSE) of 0.395.

In conclusion, this research has successfully outlined a robust approach for predicting three sorghum plant growth parameters with high accuracy using the MANFIS model and optimal membership function selection. These findings hold significant potential for advancing agriculture and aiding stakeholders in making informed decisions in sorghum cultivation. Furthermore, this study also contributes to developing ANFIS modelling techniques in the agricultural context. The optimal membership function selection applied in this research can serve as a guideline for future similar studies in predicting sorghum plant growth and other agricultural research endeavours. In future research, researchers can incorporate additional environmental factors like climate and rainfall for predicting plant growth. Additionally, they can explore other machine learning models to compare their prediction accuracy.

ACKNOWLEDGMENT

Authors are incredibly grateful to the University of Sriwijaya that provided research funding by professional competition under contract No 0127.18/UN9/SB3.LP2M.PT/202l.

REFERENCES

[1] BPS Indonesia, ‘Rice Harvest Area, Production and Productivity by Province’, Statistics Indonesia (BPS), 2022. https://www.archive.bps.go.id/indicator/53/1498/1/luas-panen-produksi-dan-produktivitas-padi-menurut-provinsi.html (accessed Mar. 12, 2022).

[2] D. Budianta, A. Napoleon, A. Paripurna, and E. Ermatita, ‘Growth and production of soybean (Glycine max (L.) Merill) with different fertilizer strategies in a tidal soil from South Sumatra, Indonesia’, Spanish J. Soil Sci., vol. 9, no. 1, pp. 54–62, Mar. 2019, doi: 10.3232/SJSS.2019.V9.N1.04.

[3] M. Sirappa, ‘Prospect of sorghum development in Indonesia as an alternative commodity for food, feed, and industry’, J. Litbang. Pert., vol. 22, no. 4, pp. 133–140, 2003.

[4] S. Sajimin, N. D. Purwantari, . S., and . S., ‘Evaluation on performance of some Sorghum bicolor cultivars as forage resources in the dry land with dry climate’, J. Ilmu Ternak dan Vet., vol. 22, no. 3, p. 135, 2018, doi: 10.14334/jitv.v22i3.1611.

[5] S. Muhrizal, J. Shamshuddin, I. Fauziah, and M. A. H. Husni, ‘Changes in iron-poor acid sulfate soil upon submergence’, Geoderma, vol. 131, no. 1–2, pp. 110–122, 2006, doi: 10.1016/j.geoderma.2005.03.006.

[6] M. Kawahigashi, N. Minh Do, V. B. Nguyen, and H. Sumida, ‘Effect of land developmental process on soil solution chemistry in acid sulfate soils distributed in the Mekong Delta, Vietnam’, Soil Sci. Plant Nutr., vol. 54, no. 3, pp. 342–352, 2008, doi: 10.1111/j.1747-0765.2008.00256.x.

[7] A. Wijanarko and A. Taufiq, ‘Effect of lime application on soil properties and soybean yield on tidal land’, Agrivita, vol. 38, no. 1, pp. 14–23, 2016, doi: 10.17503/agrivita.v38i1.683.

[8] M. Anda and D. Subardja, ‘Assessing soil properties and tidal behaviors as a strategy to avoid environmental degradation in developing new paddy fields in tidal areas’, Agric. Ecosyst. Environ., vol. 181, pp. 90–100, 2013, doi: 10.1016/j.agee.2013.09.016.

[9] Mukhlis, Y. Lestari, M. P. Yufdy, and F. Razie, ‘Effectiveness of biofertilizer formula on soil chemical properties and shallot productivity in tidal swamp land’, IOP Conf. Ser. Earth Environ. Sci., vol. 648, no. 1, 2021, doi: 10.1088/1755-1315/648/1/012158.

[10] M. Behdarnejad, H. Piri, and M. Delbari, ‘The Effect of Combined Use of Vermicompost and Poultry Manure on the Growth and Yield of Cucumber Plants in Different Conditions of Deficit Irrigation’, Water Soil, vol. 37, no. 2, pp. 237–259, 2023, doi: 10.22067/jsw.2023.79296.1215.

[11] H. Sayğı, ‘Effects of Organic Fertilizer Application on Strawberry (Fragaria vesca L.) Cultivation’, Agronomy, vol. 12, no. 5. 2022, doi: 10.3390/agronomy12051233.

[12] A. Masjedi, N. R. Carpenter, M. M. Crawford, and M. R. Tuinstra, ‘Prediction of Sorghum Biomass Using Uav Time Series Data and Recurrent Neural Networks’, in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 2695–2702, doi: 10.1109/CVPRW.2019.00327.

[13] S. Ghosal et al., ‘A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting.’, Plant phenomics (Washington, D.C.), vol. 2019, p. 1525874, 2019, doi: 10.34133/2019/1525874.

[14] S. Varela, T. Pederson, C. J. Bernacchi, and A. D. B. Leakey, ‘Understanding Growth Dynamics and Yield Prediction of Sorghum Using High Temporal Resolution UAV Imagery Time Series and Machine Learning’, Remote Sensing, vol. 13, no. 9. 2021, doi: 10.3390/rs13091763.

[15] K. H. Suradiradja, I. S. Sitanggang, L. Abdullah, and I. Hermadi, ‘Estimation of Harvest Time of Forage Sorghum (Sorghum Bicolor) CV. Samurai-2 Using Decision Tree Algorithm’, Trop. Anim. Sci. J., vol. 45, no. 4, pp. 436–442, Dec. 2022, doi: 10.5398/tasj.2022.45.4.436.

[16] A. Crane-Droesch, ‘Machine learning methods for crop yield prediction and climate change impact assessment in agriculture’, Environ. Res. Lett., vol. 13, no. 11, 2018, doi: 10.1088/1748-9326/aae159.

[17] P. Filippi et al., ‘An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning’, Precis. Agric., vol. 20, no. 5, pp. 1015–1029, 2019, doi: 10.1007/s11119-018-09628-4.

[18] H. F. Assous, H. AL-Najjar, N. Al-Rousan, and D. AL-Najjar, ‘Developing a Sustainable Machine Learning Model to Predict Crop Yield in the Gulf Countries’, Sustainability, vol. 15, no. 12. 2023, doi: 10.3390/su15129392.

[19] M. Ishak, M. S. Rahaman, and T. Mahmud, ‘FarmEasy: An Intelligent Platform to Empower Crops Prediction and Crops Marketing’, in 2021 13th International Conference on Information & Communication Technology and System (ICTS), 2021, pp. 224–229, doi: 10.1109/ICTS52701.2021.9608436.

[20] X. Xu et al., ‘Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China’, Ecol. Indic., vol. 101, no. July 2018, pp. 943–953, 2019, doi: 10.1016/j.ecolind.2019.01.059.

[21] Y. Su, H. Xu, and L. Yan, ‘Support vector machine-based open crop model (SBOCM): Case of rice production in China’, Saudi J. Biol. Sci., vol. 24, no. 3, pp. 537–547, 2017, doi: https://doi.org/10.1016/j.sjbs.2017.01.024.

[22] T. Wani, N. Dhas, S. Sasane, K. Nikam, and D. Abin, ‘Soil pH Prediction Using Machine Learning Classifiers and Color Spaces BT - Machine Learning for Predictive Analysis’, 2021, pp. 95–105.

[23] A. S. Petakar, ‘Location Based Prediction Of Crops , Analysing The Yield And Market Demand Using R-Forest and MLR’, vol. 5, no. 1, pp. 507–512, 2020.

[24] J. G. N. Zannou and V. R. Houndji, ‘Sorghum Yield Prediction using Machine Learning’, in 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART), 2019, pp. 1–4, doi: 10.1109/BIOSMART.2019.8734219.

[25] E. O. Oke et al., ‘Techno-Economic Analysis and Neuro-Fuzzy Production Rate Prediction of Sorghum (Sorghum bicolor) Leaf Shealth Colourant Extract Production’, Agric. Res., vol. 11, no. 3, pp. 579–589, 2022, doi: 10.1007/s40003-021-00596-2.

[26] Z. Lin and W. Guo, ‘Sorghum Panicle Detection and Counting Using Unmanned Aerial System Images and Deep Learning’, Front. Plant Sci., vol. 11, no. September, pp. 1–13, 2020, doi: 10.3389/fpls.2020.534853.

[27] I. Grijalva, B. J. Spiesman, and B. McCornack, ‘Computer vision model for sorghum aphid detection using deep learning’, J. Agric. Food Res., vol. 13, p. 100652, 2023, doi: https://doi.org/10.1016/j.jafr.2023.100652.

[28] H. Li, P. Wang, and C. Huang, ‘Comparison of Deep Learning Methods for Detecting and Counting Sorghum Heads in UAV Imagery’, Remote Sensing, vol. 14, no. 13. 2022, doi: 10.3390/rs14133143.

[29] T. W. Septiarini and S. Musikasuwan, ‘Investigating the performance of ANFIS model to predict the hourly temperature in Pattani, Thailand’, J. Phys. Conf. Ser., vol. 1097, no. 1, p. 12085, 2018, doi: 10.1088/1742-6596/1097/1/012085.

[30] M. Şahin and R. Erol, ‘A Comparative Study of Neural Networks and ANFIS for Forecasting Attendance Rate of Soccer Games’, Mathematical and Computational Applications, vol. 22, no. 4. 2017, doi: 10.3390/mca22040043.

[31] M. Rezaei, A. Rohani, and S. S. Lawson, ‘Using an Adaptive Neuro-fuzzy Interface System (ANFIS) to Estimate Walnut Kernel Quality and Percentage from the Morphological Features of Leaves and Nuts’, Erwerbs-Obstbau, vol. 64, no. 4, pp. 611–620, 2022, doi: 10.1007/s10341-022-00706-6.

[32] V. R. Phate, R. Malmathanraj, and P. Palanisamy, ‘Clustered ANFIS weighing models for sweet lime (Citrus limetta) using computer vision system’, J. Food Process Eng., vol. 42, no. 6, p. e13160, Oct. 2019, doi: https://doi.org/10.1111/jfpe.13160.

[33] Tarno, A. Rusgiyono, and Sugito, ‘Adaptive Neuro Fuzzy Inference System (ANFIS) approach for modeling paddy production data in Central Java’, J. Phys. Conf. Ser., vol. 1217, no. 1, p. 12083, 2019, doi: 10.1088/1742-6596/1217/1/012083.

[34] M. A. Raharja, I. D. M. B. A. Darmawan, D. P. E. Nilakusumawati, and I. W. Supriana, ‘Analysis of membership function in implementation of adaptive neuro fuzzy inference system (ANFIS) method for inflation prediction’, J. Phys. Conf. Ser., vol. 1722, no. 1, 2021, doi: 10.1088/1742-6596/1722/1/012005.

[35] N. Talpur, M. N. M. Salleh, and K. Hussain, ‘An investigation of membership functions on performance of ANFIS for solving classification problems’, IOP Conf. Ser. Mater. Sci. Eng., vol. 226, no. 1, 2017, doi: 10.1088/1757-899X/226/1/012103.

[36] M. Babanezhad, A. Masoumian, A. T. Nakhjiri, A. Marjani, and S. Shirazian, ‘Influence of number of membership functions on prediction of membrane systems using adaptive network based fuzzy inference system (ANFIS)’, Sci. Rep., vol. 10, no. 1, pp. 1–20, 2020, doi: 10.1038/s41598-020-73175-0.

[37] N. Gupta and R. S. Sharma, ‘A Comparative Study of ANFIS Membership Function to Predict ERP User Satisfaction using ANN and MLRA’, Int. J. Comput. Appl., vol. 105, no. 5, pp. 11–15, 2014.

[38] O. Adil, A. Ali, M. Ali, A. Y. Ali, and B. S. Sumait, ‘Comparison between the Effects of Different Types of Membership Functions on Fuzzy Logic Controller Performance’, Int. J. Emerg. Eng. Res. Technol., vol. 3, no. April, p. 76, 2015, [Online]. Available: https://www.researchgate.net/publication/282506091.

[39] M. H. Azam, M. H. Hasan, S. J. Abdul Kadir, and S. Hassan, ‘Prediction of Sunspots using Fuzzy Logic: A Triangular Membership Function-based Fuzzy C-Means Approach’, Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 2, pp. 357–362, 2021, doi: 10.14569/IJACSA.2021.0120245.

[40] S. Amid and T. Mesri Gundoshmian, ‘Prediction of output energies for broiler production using linear regression, ANN (MLP, RBF), and ANFIS models’, Environ. Prog. Sustain. Energy, vol. 36, no. 2, pp. 577–585, Mar. 2017, doi: https://doi.org/10.1002/ep.12448.

[41] M. Alizadeh, M. Gharakhani, E. Fotoohi, and R. Rada, ‘Design and analysis of experiments in ANFIS modeling for stock price prediction’, Int. J. Ind. Eng. Comput., vol. 2, no. 2, pp. 409–418, 2011, doi: 10.5267/j.ijiec.2011.01.001.

[42] J. S. R. Jang, C. T. Sun, and E. Mizutani, ‘Neuro-Fuzzy and Soft Computing-A Computational Approach to Learning and Machine Intelligence [Book Review]’, IEEE Trans. Automat. Contr., vol. 42, no. 10, pp. 1482–1484, 2005, doi: 10.1109/tac.1997.633847.

[43] R. Ata and Y. Kocyigit, ‘An adaptive neuro-fuzzy inference system approach for prediction of tip speed ratio in wind turbines’, Expert Syst. Appl., vol. 37, no. 7, pp. 5454–5460, 2010, doi: 10.1016/j.eswa.2010.02.068.

[44] D. Chicco, M. J. Warrens, and G. Jurman, ‘The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation’, PeerJ Comput. Sci., vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

A Hybrid Double Encryption Approach for Enhanced Cloud Data Security in Post-Quantum Cryptography

Manjushree C V[1], Nandakumar A N[2]
Information Science and Engineering, Vemana Institute of Technology (VTU University), Bangalore, India[1]
Computer Science and Engineering, VTU/City Engineering College, Bangalore, India[2]

***Abstract*—Quantum computers and research on quantum computers are increasing due to the efficiency and speed required for critical applications. This scenario also kindles the vitality of data protection needed against threats from quantum computers. Research in post-quantum threats is very minimal so far, but it is much needed to protect the enormous data stored in the cloud for healthcare, governmental, or any crucial data. This research work presents an advanced hybrid double encryption approach for cloud data security based on Post-Quantum Cryptography (PQC) to ensure the restriction of unauthorized access. The suggested approach combines the benefits of the NTRU encryption and AES encryption algorithms and works in hybrid mode, offering strong security while resolving issues with real-time performance and cost-efficiency. A streamlined key management system is set together to improve real-time processing, significantly reducing encryption and decryption delay times. Moreover, NTRU Encrypt dynamic parameter selection, which adapts security parameters based on data sensitivity, maintains accurate information and security. In addition to addressing real-time performance and data security, an innovative development in this method is known as Quantum-Adaptive Stream Flow Encryption (QASFE), which enables secure data sharing and collaborative working within a quantum-resistant framework. This innovative feature enhances data accessibility while maintaining the highest level of security. In the era of post-quantum cryptography, our multifactor authentication technique, integrating double encryption and QASFE, is a proactive and flexible solution for securing cloud data, and protecting data security and privacy against emerging threats.**

Keywords—Cloud data security; double encryption; Post-Quantum Cryptography (PQC); NTRU Encrypt; AES Encryption

I. Introduction

Data security in the cloud has become essential in the present digital landscape as organizations depend upon increasing amounts on cloud-based services to safeguard, analyze, and manage their data. The cloud certainly radiates in three areas: scalability, accessibility, and cost-effectiveness. Data breaches, insider threats, and the ever-changing character of attacks are just a few of the specific security challenges that it additionally raises. Cloud computing technology has rendered it possible for many firms to design and deploy software with greater efficacy and effectiveness in the cloud, enabling savings on the expenses of purchasing and maintaining the infrastructure [1]. The term "cloud computing" refers to the idea of rapid, on-demand network access to a pool of programmable computing resources (including networks, storage devices, servers, apps, and services). It enables these assets to be delivered and released rapidly, with the least amount of managerial work and service provider engagement. Since security is the primary concern in cloud computing, some users may find themselves comfortable transmitting information via the cloud. Cloud computing specialists have created certain secret keys for account security and use encryption methods to safeguard cloud servers [2, 3, 5]. To safeguard the data from threatening attacks, specific encryption techniques might be employed. The cloud security posture is strengthened through frequent security assessments, incident response plans, and adherence to industry-specific laws and standards.

This research study was established by combining different security strategies to achieve the objective of cloud data security. Combining these techniques creates an obstacle against security issues, which have been preventing the cloud's acceptance and effective functioning. The field of cryptography has evolved in reaction to this impending quantum threat, which has contributed to the conception of Post-Quantum Cryptography (PQC). The major goal of PQC is to provide encryption techniques and cryptographic primitives that are resistant to quantum attacks. Attacks are computationally expensive and consequently ineffective as an outcome. These basic hypotheses, on which conventional stocks depend, are no longer valid with the development of post-quantum computing [4]. A shield against security issues that have frequently prevented the cloud from functioning and growing effectively is created by combining these techniques. Data security in the cloud utilizes a sophisticated strategy that combines innovative technology, widely accepted practices, and stringent regulations to safeguard data from unauthorized access, data breaches [5], and other cyber consequences of these challenges. One such cryptographic method combines NTRU Encrypt and AES encryption, resulting in an effective Double Encryption method. NTRU Encrypt, a lattice-based encryption system developed to withstand the cryptographic flaws that quantum computers can eventually exploit, is one of the notable PQC algorithms in this field. NTRU Encrypt, an effective encryption algorithm constructed on the mathematical underpinnings of lattice-based cryptography, is at the core of PQC's promise [6].

The Advanced Encryption Standard (AES) [7, 8] has evolved as an essential component of contemporary cryptographic methods in the pursuit of secure data transmission and storage. AES has become a cornerstone of data security in several industries owing to its dependability and effectiveness in protecting sensitive data. As the secondary encryption layer, AES is essential to post-quantum cryptography and the idea of enhanced double encryption. It increases the quantum resistance offered by NTRU Encrypt and provides an additional level of security while also improving the impact and effectiveness of data protection as an entire.

Fig. 1 shows the essential steps in NTRU Encrypt and AES-based double encryption. It offers a strong security framework that makes use of the advantages of both encryption techniques to safeguard data in a way that is quantum-resistant while preserving computing effectiveness.

The following are the main contributions of this work:

- Post-Quantum Security: The primary contribution consists of implementing the post-quantum cryptography method, particularly NTRU Encrypt, to improve data security. It incorporates NTRU Encrypt, which is made resistant to quantum attacks, to address the immediate challenge that quantum computing presents to conventional cryptographic techniques.
- Double Encryption Strategy: Introduces a strong double encryption technique called the Double Encryption Strategy, which combines the advantages of the NTRU Encrypt and AES encryption techniques. Sensitive data is further protected with our dual-layered encryption strategy's extra degree of security.
- Key Management: Effective key management is essential for any encryption scheme, and the paper recognizes this by highlighting key generation as a significant contribution. It describes how to generate and manage cryptographic keys securely, which is crucial for the overall safety of the encryption system.

Quantum-adaptive Security: The primary benefit of this paper is the creation of quantum-adaptive security in the context of post-quantum cryptography. QASFE provides an unparalleled degree of adaptability and security by dynamically modifying encryption procedures based on the quantum threat level and data relevance.

This paper is structured as follows: Section II provides an overview of the literature review for the security of data in recent cloud computing research. Section III, a methodology is proposed to secure data prevention in cloud environments. Section IV presents the results and accompanying discussion, and Section V concludes this paper.

II. Literature Survey

A. Related Works

This section provides several recent cloud computing experiments. Even though many research efforts have focused on the security risks with cloud computing cryptographic techniques. Some researchers suggest employing innovative methods in addition to the strategies defined above to improve cloud computing security.

The literature review, which was conducted is provided in tabulated form (see Table I), provides an overview of the available work, and has established a ground for the proposed work. In this section, the approaches proposed by various researchers about the issue have been addressed for the recognition of the problem and understanding.

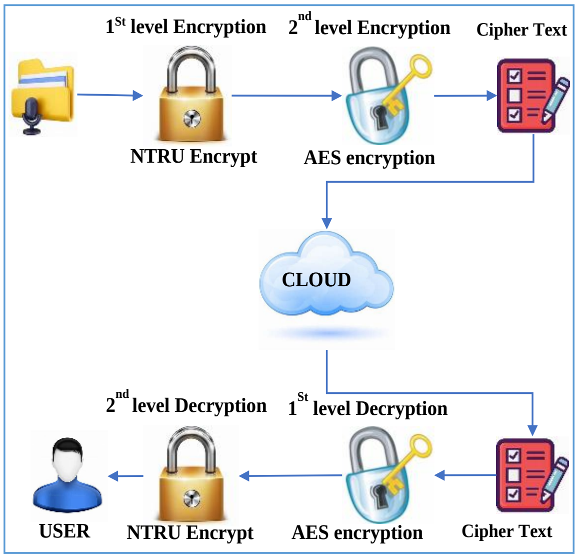


Fig. 1. Basic architecture for double encryption.

TABLE I. Analysis of Related Works

Reference	Focus	Description
Singh, Prabhdeep, and Ashish Kumar Pandey, et al. [9]	Data Security in Cloud	This study offers a thorough analysis of the literature on data encryption, data concealment, and data protection challenges, as well as solutions for cloud data storage. The effectiveness of each strategy is then contrasted based on its features, advantages, and drawbacks.
A. Malviya, R. K. Dwivedi, et al. [10]	Comparison of container orchestration tool	This research will assist professionals in determining if they require an orchestrator that is connected to a certain technology or one that offers an independent solution. Data integrity, availability, and secrecy are all referred to be forms of security, though. As long as they don't provide or guarantee this for this paperwork, there could be major issues.
Achar, Sandesh, et al.,[11]	Infrastructure as Code (IAC) across multiple clouds	This research indicates that there are a lot of challenges with managing the infrastructure provisioning of enterprise SaaS applications, including configuration drift and the variety of cloud providers. Without compromising on stability or quality, IAC makes it possible to roll out novel application architecture versions rapidly. Anytime a cloud host isn't accessible, there should be a warning period before the services restart.
Singh, A. K et al. [12]	Protecting the cloud data from unauthorized access	The researchers' proposed method complicates cryptanalysis by changing the plain text alphabet's position using prime numbers and then performing the hybrid RSA algorithm, which makes it more challenging to factorize the variable used in key creation. The main problem with employing is excessive resource utilization, which eventually results in higher costs and longer wait times.
Nejatollaji et al. [13]	Resistance against the dangers of quantum computing	This study examines trends in lattice-based cryptographic methods, including recent fundamental concepts for the use of lattices in computer security, difficulties in implementing them in software and hardware, and new

	in post-quantum cryptography.	requirements for their adoption. The survey also offered insightful ideas that would enable the reader to concentrate on the mechanics of the computation ultimately required for mapping schemes on already-existing equipment or generating a scheme in part or in full on specialized hardware.
Dutta, Aritra, et al., [14]	Multi-cloud storage's security for preventing unauthorized people from accessing data	According to this research, the business decides to utilize a username and password along with the NTRU encryption technique to encrypt the data to increase security. By doing this, it is possible to ensure that even if an attacker obtains access to the encrypted data, they will not be able to decrypt it without the correct password.
Harjito, Bambang, et al., [15]	Threats to the security of digital information	This study's findings demonstrate that the NTRU algorithm's key generation and encryption times are quicker than those of the RSA technique. The NTRU approach is better advised for cloud storage security because of its higher level of resiliency security. The NTRU algorithm, on the other hand, employs a lattice-based strategy with strong key selection and is also difficult to solve.
Costa, Bruno, et al. [16]	Security in the quantum universal composability	In this work, they propose randomized variants of two well-known oblivious transfer protocols, one quantum, and the other post-quantum, with ring training and an error assumption. The accuracy is sacrificed in preference to cost savings with these biometric authentication systems that have been reduced. This paper highlighted contemporary technological technologies like cloud computing, it is now possible to access data from any location.
Tarannum, Ayesha, et al. [17]	Data security method using multi-modal biometric sensing and authentication	To enable robust data integrity verification and data security in distributed applications, a new integrity computational algorithm and an encryption method have been developed in this work. As a result, each user receives a unique set of variable keys during the encryption process for additional security. This paradigm, however, is unable to provide the needed results in real time for multi-modal biometric authentication.

B. Motivation of the Proposed Research

As we've seen, a lot of research has been done on encryption, and in the majority of post-quantum cryptography research, the ranking operations are performed only on the platform side. Therefore, it is necessary to propose the PQC with NTRU-AES Encrypt.

III. Proposed Methodology

The adoption of strong post-quantum cryptographic solutions receives priority in the strategy due to the impending threat posed by quantum technology. The introduction of a revolutionary Double Encryption methodology, which combines the benefits of the NTRU Encrypt and AES encryption methods while including novel elements to improve data security and accessibility, is a key component of this strategy. Furthermore, it enhances key management processes to improve real-time performance, reducing encryption and decryption lag and enabling safe, on-the-fly data transfer and retrieval. This improved Double Encryption technique uses NTRU Encrypt and AES encryption for key generation to provide a diverse defense against potential quantum and classical attacks, which is necessary to combat the growing threat of quantum attacks. In addition, the methodology provides an innovative Quantum-Adaptive Stream Flow Encryption (QASFE) component that addresses temporal performance measurements and creates a simplified key management system to reduce latencies. This method offers a thorough and effective method for protecting data in the post-quantum era of cryptography while working to significantly reduce the time needed for encryption and decryption, ensuring security and real-time data accessibility. It does this by dynamically adjusting security parameters and adapting key sizes to data priority. An overview of a proposed methodology for improving data security in cloud systems is provided in Fig. 2. Strong Post-Quantum Cryptographic solutions are required as quantum computers approach. To address this issue, we provide a new Double Encryption technique that combines the benefits of NTRU Encrypt and AES encryption besides incorporating innovative components that improve data security and accessibility.

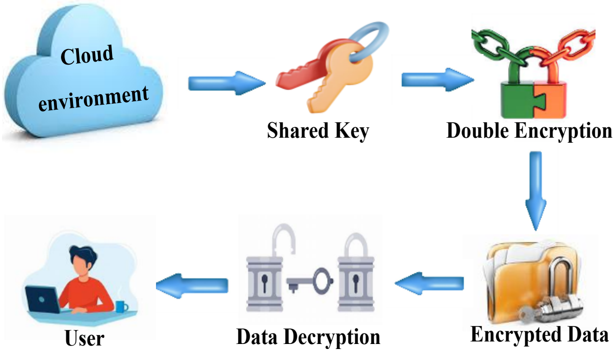


Fig. 2. Overview of proposed methodology for enhanced data security.

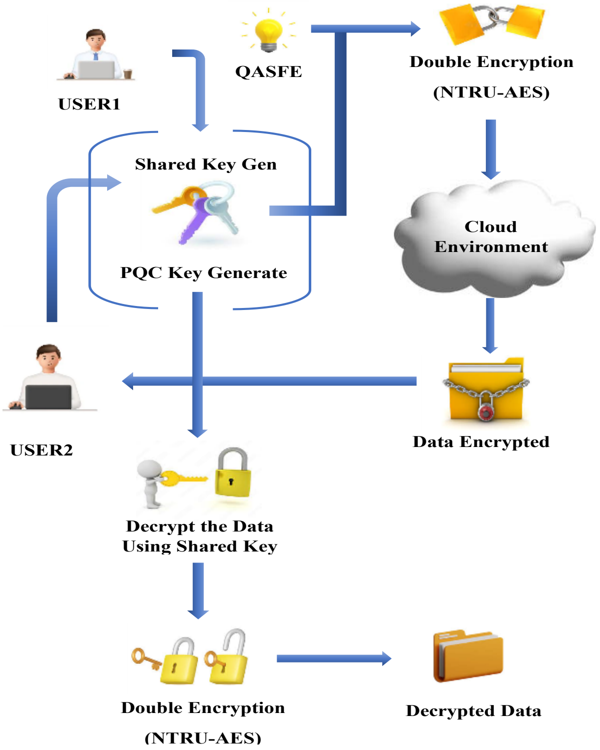


Fig. 3. Proposed architecture for double encryption data security.

To protect against quantum attacks, our enhanced Double Encryption method originates by using NTRU Encrypt and AES encryption for key generation. We optimize the key management procedure, though, to address queries regarding real-time performance. We want to decrease the latency of encryption and decryption by optimizing key generation and management. This would enable secure real-time data transfer and retrieval.

The combination of various cryptographic methods provides an effective defense against potential quantum attacks and classical attacks, improving the security posture generally. We aim to reduce the time needed for encryption and decryption by the optimization of the key generation procedure, enabling real-time data transfer and retrieval operations without sacrificing security. We add a unique component to our enhanced Double Encryption approach called Quantum-Adaptive Stream Flow Encryption (QASFE), which addresses the time performance measures. Due to the delay in period of time, our approach proposes a streamlined key management system that effectively creates and manages encryption keys to address this problem. Our approach, which relies on NTRU Encrypt and AES key generation, allows us to utilize smaller key sizes for high-priority data and larger key sizes for low-priority data to improve security. Our enhanced Double Encryption technique resolve the problems with real-time efficiency and also incorporates the revolutionary notion of Quantum-Adaptive Stream Flow Encryption (QASFE). We offer a comprehensive and effective method for protecting data in the era of post-quantum cryptography by enhancing key management, dynamically modifying security parameters, and implementing QASFE. Below Fig. 3, the proposed Double Encryption Data Security system is described. It combines NTRU Encrypt and AES encryption with a focus on quantum resistance, time efficiency, and data security in cloud environments. A strong and flexible security framework for the post-quantum generation is created by combining NTRU Encrypt, AES encryption, and the QASFE component.

A. *NTRU Encrypt - AES Encryption*

An asymmetric key (public - private key) and symmetric key (public key) cryptosystem termed the NTRU-AES encryption method is parameterized by three integers: (I, l, m), where p is greater than q, GCD (p, q) = 1, and N is prime. Polynomial rings P, Rp, and Rq should be considered. x^N-1 is an ideal in Z[x], that is, it contains all polynomials in H[x] that can be represented as multiples of a^I-1.

$$P = \frac{H[x]}{a^I - 1}, R_p = \frac{\frac{Z}{P} H[x]}{a^I - 1}, R_q = \frac{\frac{Z}{P} H[x]}{a^I - 1}$$

The equation for the product of two polynomials a(x), b(x) ϵ R is given as

a(x) * b(x) = c(x)

$$C_K = \sum_{i=j=k \ (mod\ N)} a_i \, b_{k-i}$$

Here;

The variable on which the phrase C_K depends is K, which most likely denotes the result or output of this function.

The term inside the sum, $i = j = k\ (mod\ N)$, designates the range of values that the index variable i can take. For each value of i that meets the congruence criterion, the product of two sequences, $a_i \ b_{k-i}$, is contained within the summation.

To indicate any positive integers w1 and w2, let τ (w1, w2) denotes the set of ternary polynomials given by,

$$\tau \text{ (w1, w2) } = \left\{ z(x) \ \in \right.$$

$$R \left. \begin{matrix} The\ d1\ coefficients\ of\ z(x)\ are\ equal\ to\ 1, \\ The\ d2\ coefficients\ of\ z(x) are\ equal\ to\ -1, \\ All\ other\ coefficients\ of\ z(x)\ are\ equal\ to\ 0 \end{matrix} \right\}$$

The polynomial z(x) with particular coefficient configuration composed the set τ (w1, w2). It includes polynomials with w1 coefficients of 1, w2 coefficients of -1, and w0 coefficients of 0. Based on these coefficient requirements, a set of polynomials can be defined utilizing this type of notation.

Key Generation: Implementing "NTRU-AES," a synergistic cryptographic technique that combines the mathematical efficacy of NTRU Encrypt for the generation of public and private keys with the dependability of AES for the generation of symmetric keys to produce a hybrid encryption system that provides unrivaled security in a quantum-resistant framework.

Choose private f(x) ϵ τ (w + 1, w)

Choose private v(x) ϵ τ (w, w).

- Generate f_Q (x) = f^{-1}(x) in R_q and f_p (x) = f^{-1}(x) in R_p
- Generate h (x) = f_q (x) * v(x) in R_q

Users' public keys are represented by polynomial h(x). Pair (f(x), f_p (x)) is the corresponding private key. The user can store only f(x) and regenerate f_p (x) from it. The encryption processes for plaintext m(x) ϵ R_p are indicated in the following phases.

Encryption: An essential part of encryption, and the security of the encryption protocol depends on the complexity of several mathematical challenges, such as the problem of encoding to obtain the cipher text message and the challenge of factoring the modulus employed in the polynomial ring.

- Select a random transitory key r(x) ϵ τ (w, w).

To perform this, employ an encoding technique that converts the message's plaintext to the ciphertext of the polynomial ring.

- Compute the ciphertext e(x) = pr (x) * h(x) + m(x) mod q.

The secure ciphertext matching the initial polynomial equation will be the result of adhering to NTRU and AES encryption (if used). The matching decryption method (which undoes the stages of encryption) can be used to decrypt this ciphertext when required. It can be safely stored in the cloud.

Decryption: Data is transformed into an unreadable format using encryption algorithms and keys when it is encrypted before it is stored in the cloud. Authorized users or programs must decrypt the data to access and use it.

- Compute a(x) = f(x) * e(x) mod q
- Center lift a(x) to a(x) ϵ R
- Compute m = f_p (x) * a (x) mod p

Notably, the polynomial multiplication in the final decryption phase is omitted by selecting f(x) = 1+p f_1 (x), where f1 R in the key generation step as f_p = 1 mod p.

The overall result, m, is the message that has been decrypted. It is calculated by multiplying the result of f_p (x) by the modified a(x), then applying modulo p.

The precise perception and relevance of these operations will depend on the exact meanings of f(x), e(x), f_p (x), q, and p as well as the context in which this decryption procedure is utilized. The specifics of this procedure may vary based on the particular mathematical problem or cryptographic technique being used, although it is frequently observed in certain mathematical procedures and systems.

B. *QASFE Framework*

The real-time performance and data security in the cloud are enhanced by using the sophisticated mathematical framework of Quantum-Adaptive Stream Flow Encryption (QASFE). Lattice-based cryptography, a field of mathematics that involves solving difficult problems involving lattices, like the Shortest Vector Problem (SVP) and Learning with Errors (LWE), is the foundational idea of QASFE. To improve data security and manage real-time performance issues in cloud-based scenarios, lattice problems are exploited for their inherent computational complexity.

Lattice Basis: A group of vectors that are linearly separate that span a lattice, b1, b2, ..., bn, define the lattice as L:

$$\text{Lattice L = Span (b1, b2, ..., bn)}$$

$$L = \{x \mid x = \sum(ai * bi) \text{ for } ai \in Z,\ 1 \leq i \leq n\}$$

In order to enhance the real-time performance of lattice-based cryptography, the aforementioned formula emphasizes improving the efficiency of lattice-based encryption and decryption, which is crucial for real-time performance:

$$O\ (D^3 * N * log^2\ (BS)) \qquad (1)$$

Here:

L represents the lattice

x represents any vector that belongs to the lattice L, the symbol $\sum$ denotes summation,

ai represents integer coefficients.

bi represents basis vectors of the lattice.

Z represents the set of integers.

D represents the dimension of the lattice,

N is the number of lattice vectors.

BS is a bound on the size of the coefficients.

To optimize these parameters in Eq. (1) to enhance the real-time performance of lattice-based cryptography, it might be required to decrease the dimension, restrict the number of lattice vectors, or adopt lower parameter bounds. With the aid of these optimizations, encryption and decryption processes will be quicker and more effective.

Shortest Vector Problem (SVP): In a given lattice L, the Shortest Vector Problem (SVP) attempts to identify the shortest non-zero matrix as follows:

$$SVP(L) = \min \{\|v\| \mid v \in L, v \neq 0\}$$

$$TP = (2^d) / (SVP\ (L)^c) * \log2(N) \qquad (2)$$

where,

SVP(L) represents the lattice's average shortest vector length L.

min {||v|| | v ϵ L, v $\neq$ 0} signifies the search for the minimum Euclidean norm (length) of a non-zero vector v within the lattice L, 2^d represents the dimensionality of the lattice's parabolic expansion in computational complexity.

log2(N) represents the performance impact of basis size is expressed by the logarithm of the number of lattice basis vectors with base 2,

Eq. (2) describes the dimension of the lattice, the average shortest vector length, the number of basis vectors, and the computing difficulty of solving the SVP in a lattice-based cryptographic system.

Learning with Errors (LWE): LWE utilizes an error e from a particular distribution D in conjunction with a vector s from Z q^n.

Finally, a set of noisy linear equations is computed:

a = (a1, a2, ..., an) ∈ Z q^n

b = <a, s> + e mod q

a ← Uniform polynomial in R q.

e ← Error polynomial in R q.

b = a*s + e

ciphertext = (a, b)

where,

'a' is a vector containing 'n' integers from the ring Z q.

's' is a secret vector (usually known only to the recipient of the ciphertext).

'e' is a noise term introduced to enhance security.

'<a, s> ' represents the dot product of 'a' and 's'.

The result 'b' is computed as the dot product plus the noise term, all taken modulo q.

ciphertext = (a, b) represents the result of the LWE encryption process.

This encryption procedure adds some noise ('e') to the ciphertext in the context of LWE, rendering it extremely difficult for attackers to extract the private vector's' from the ciphertext. Even with the use of innovative quantum techniques, it is difficult for attackers to separate the original plaintext from the ciphertext due to the inclusion of erroneous phrases in the encryption process.

From QASFE, the mathematical foundations of lattices, which include the Shortest Vector Problem (SVP) and Learning with Errors (LWE), enhance their resistance to quantum attacks. A scenario where data remains private and secure in the cloud has become possible due to the adoption of such secure cryptographic techniques, which become increasing essential as quantum computing develops.

IV. Results and Discussion

The proposed Double Encryption methodology has undergone thorough evaluation and testing. It combines NTRU Encrypt and AES encryption and is enhanced by the innovative QASFE component. In this instance, we examined the data security strategy, focusing on crucial performance indicators like entropy and throughput. The implementation was performed using Python programing language.

The table illustrates various settings for entropy, and throughput. The comparison of several algorithms is shown in Table II along with relevant metrics including the proposed method's entropy and encryption and decryption time.

TABLE II. Comparing several Techniques based on Variables Like Entropy, Encryption Time, and Decryption Time

Methods	Average entropy per byte	Throughput (kb/sec)
AES	3.840	192.00
RSA	3.095	100.06
BLOWFISH	3.892	197.20
PROPOSED	7.440	237.12

A. Entropy

The term "entropy" refers to a gauge of the random or unpredictability of data in the setting of data security and cryptography. Higher entropy values imply that the method of encryption is more advanced and secure since it becomes more difficult for a competitor to predict or decode the encrypted data. The average entropy per byte for several encryption techniques is shown in Fig. 4, providing insight into the level of data protection for every strategy. Particularly, the proposed method far outperforms the competition, sporting high average entropy per byte of 7.44, indicating a significant improvement in data security compared to the conventional encryption techniques.

B. Throughput

It is a metric for gauging the way effectively processors respond to an algorithm. The volume of data transferred between users serves as a measure of it. The file uploading phase must have a high throughput that allows for the transfer of all data. When uploading files, the proposed approach displays the highest throughput.

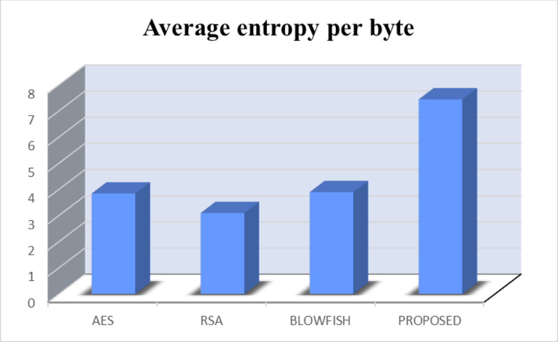


Fig. 4. Average entropy values of proposed model with existing AES, RSA and Blowfish algorithm. (x-axis: unit per byte, y-axis: Various encryption techniques).

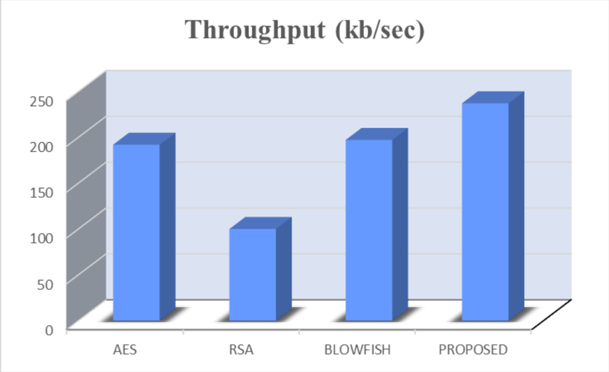


Fig. 5. Throughput values of the proposed model compared with existing AES, RSA, and Blowfish algorithm (x-axis: kB/s, y-axis: Various encryption techniques).

Fig. 5 compares the throughput of various cryptographic techniques, such as AES, RSA, and BLOWFISH, in kilobytes per second (kb/sec), with the proposed method. The proposed technique shines apart by possessing a higher throughput of 237.12 kb/sec, indicating the speed with which it processes data. These results imply that the proposed strategy demonstrates superior data processing abilities, making it an appropriate pick for applications that require both security and performance.

V. Conclusion

In conclusion, our suggested data security design offers a significant progress in the field of cloud security and data protection, specifically in considering recent developments in quantum computing. We have improved data security along with addressing critical concerns regarding real-time performance and computational prosperity by developing an innovative Double Encryption technique that combines NTRU Encrypt with AES encryption and the quantum-resistant properties of Post-Quantum Cryptography (PQC), offering an effective defense against the evolving threats and challenges of the digital era. Our core innovations, such as the unique QASFE, dynamic security parameter modifications, and improved key management, accessible the path for a data security system that is more time-sensitive, resource-effective, and resilient. This study reimagines the data protection environment by giving enterprises a strong, flexible, and quantum-resistant solution that enables them to protect their most important data while ensuring quick and secure data transfers.

REFERENCES

[1] Jones, Kofi Immanuel, and R. Suchithra. "Information Security: A Coordinated Strategy to Guarantee Data Security in Cloud Computing." International Journal of Data Informatics and Intelligent Computing 2.1 (2023): 11-31, 2023.

[2] Kulshrestha, Vartika, Seema Verma, and C. Rama Krishna. "Hybrid probabilistic triple encryption approach for data security in cloud computing." International Journal of Advanced Intelligence Paradigms 21.1-2, 158-173.

[3] Gupta, Ishu, et al. "Compendium of data security in cloud storage by applying hybridization of encryption algorithm", Institute of Electrical and Electronics. Engineers (IEEE), 2022.

[4] Alu, Esther S., Kefas Yunana, and Muhammed U. Ogah. "Secured Cloud Data Storage Encryption Using Post-Quantum Cryptography." International Journal of Advanced Research in Computer and Communication Engineering. Vol. 11, Issue 7, July 2022.

[5] Devi, B. Padmini, and S. Kannadhasan. "Preventing Data Leakage in Cloud Servers through Watermarking and Encryption Techniques." Research Square, 2023.

[6] Ukwuoma, Henry C., et al. "Quantum attack-resistant security system for cloud computing using lattice cryptography." International Journal for Information Security Research 12.1, 2022.

[7] Alemami, Yahia, et al. "Cloud data security and various cryptographic algorithms." International Journal of Electrical and Computer Engineering 13.2, 2023.

[8] Fatima, Sana, et al. "Comparative Analysis of Aes and Rsa Algorithms for Data Security in Cloud Computing." Engineering Proceedings 20.1, 2022.

[9] Singh, Prabhdeep, and Ashish Kumar Pandey. "A Review on Cloud Data Security Challenges and Existing Countermeasures in Cloud Computing." International Journal of Data Informatics and Intelligent Computing 1.2, 23-33, 2022.

[10] A. Malviya and R. K. Dwivedi, "A Comparative Analysis of Container Orchestration Tools in Cloud Computing," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 698-703, 2022.

[11] Achar, Sandesh. "Enterprise SaaS Workloads on New-Generation Infrastructure-as-Code (IaC) on Multi-Cloud Platforms." Global Disclosure of Economics and Business 10.2, 55-74, 2021.

[12] Gupta, I., Gurnani, D., Gupta, N., Singla, C., Thakral, P., & Singh, A. K., “Compendium of data security in cloud storage by applying hybridization of encryption algorithm”, Institute of Electrical and Electronics Engineers (IEEE), 2022.

[13] Nejatollahi, Hamid, et al. "Post quantum lattice-based cryptography implementations: A survey." ACM Computing Surveys (CSUR), 51.6, 1-41, 2019.

[14] Dutta, Aritra, et al. "A New Encryption Algorithm Helps to Secure the Cloud Storage." International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 16, Number 1, 2023.

[15] Harjito, Bambang, et al. "Comparative Analysis of RSA and NTRU Algorithms and Implementation in the Cloud." International Journal of Advanced Computer Science and Applications 13.3, 2022.

[16] Costa, Bruno, et al. "Randomized Oblivious Transfer for Secure Multiparty Computation in the Quantum Setting." Entropy 23.8, 1001, 2021.

[17] Tarannum, Ayesha, et al. "An efficient multi-modal biometric sensing and authentication framework for distributed applications." IEEE Sensors Journal, 20.24, 15014-15025, 2020.

Comprehensive Analysis of Topic Models for Short and Long Text Data

Astha Goyal[1], Indu Kashyap[2]
Research Scholar, Department of CSE, MRIIRS, Faridabad, India[1]
Professor, Department of CSE, MRIIRS, Faridabad, India[2]

***Abstract*—The digital age has brought significant information to the Internet through long text articles, webpages, and short text messages on social media platforms. As the information sources continue to grow, Machine Learning and Natural Language Processing techniques, including topic modeling, are employed to analyze and demystify this data. The performance of topic modeling algorithms varies significantly depending on the text data's characteristics, such as text length. This comprehensive analysis aims to compare the performance of the state-of-the-art topic models: Nonnegative Matrix Factorization (NMF), Latent Dirichlet Allocation using Variational Bayes modeling (LDA-VB), and Latent Dirichlet Allocation using Collapsed Gibbs-Sampling (LDA-CGS), over short and long text datasets. This work utilizes four datasets: Conceptual Captions and Wider Captions, image captions for short text data, and 20 Newsgroups news articles and Web of Science containing science articles for long text data. The topic models are evaluated for each dataset using internal and external evaluation metrics and are compared against a known value of topic 'K.' The internal and external evaluation metrics are the statistical metrics that assess the model's performance on classification, significance, coherence, diversity, similarity, and clustering aspects. Through comprehensive analysis and rigorous evaluation, this work illustrates the impact of text length on the choice of topic model and suggests a topic model that works for varied text length data. The experiment shows that LDA-CGS performed better than other topic models over the internal and external evaluation metrics for short and long text data.**

***Keywords*—Topic modeling; Nonnegative Matrix Factorization (NMF); Latent Dirichlet Allocation (LDA); evaluation metrics; short text mining; long text mining**

I. Introduction

Topic modeling has emerged as a powerful technique for uncovering hidden thematic structures within large volumes of textual data. It provides valuable insights by automatically identifying and extracting topics from unstructured text. It is a powerful tool for text analysis, and it has been used for various applications, including document classification, summarization, and recommendation systems. The basic idea behind topic modeling is to assume that each document in a corpus can be represented as a mixture of topics. The topics are latent variables that are not directly observed in the data. However, the topics can be inferred from the words that appear in the documents.

The performance of topic modeling algorithms is influenced by the characteristics of the data, including its length. Short text data poses challenges in finding the co-occurrence of topic patterns due to data sparsity, noise, topic imbalance, and lack of context [1]. Conversely, long text data presents computational complexity, overfitting, and interpretability issues when employing topic modeling algorithms. It is also hardly feasible to infer a unique and coherent topic from a long text because it usually contains various topics [2]. In this research domain, the foremost obstacle lies in selecting an optimal topic model that remains effective irrespective of the length of the text or that works for the specific application domain. The secondary challenge entails the fine-tuning and enhancement of the chosen topic model.

The present study aims to compare various topic modeling algorithms in the context of short and long text data, investigating their effectiveness and efficiency under different conditions. After an extensive review of existing literature in this field, LDA emerges as a top-performing model for extensive text data. At the same time, NMF excels in shorter text contexts [3], [4]. Recent attention has also turned to LDA's adaptability. One key feature that makes LDA adaptable is the use of the Bayesian framework. This means that LDA incorporates the prior knowledge about the data into the model, improving the model's performance, especially for small or noisy datasets. Therefore, this study evaluates the NMF topic model against two distinct variations of the LDA topic model: LDA employing Variational Bayes (LDA-VB) and LDA using Collapsed Gibbs Sampling (LDA-CGS). The goal is to introduce a topic modeling approach that remains effective regardless of the text length.

These models are subjected to rigorous evaluation using internal and external evaluation metrics, referencing the known number of topics 'K' to ensure unbiased results [5], [6]. Internal evaluation metrics measure the quality of the topic model itself without reference to any external data [7]. External evaluation metrics measure the quality of the topic model by evaluating its performance on a downstream task, such as text classification or document clustering [6]. The findings of these experiments illuminate how text length influences topic modeling outcomes, offering insights into selecting the most suitable topic model regardless of text length. Additionally, to ensure the robustness of the evaluation, a diverse selection of datasets encompassing both long and short text data, thereby mirroring the heterogeneity of real-world textual information sources, has been thoughtfully chosen to ensure the robustness of the evaluation. Namely, the Conceptual Captions [8] and WIDER Captions [9] datasets are selected for short text, and the 20

Newsgroups [10] and Web of Science [11] datasets are selected for long text.

The paper is structured as follows: Section II discusses the related work for topic models NMF and LDA. The experimental exploration for comparing different topic models on both short and long text data using evaluation metrics is presented in Section III. Section IV presents a comparative analysis of the obtained results. The paper is concluded in Section V with suggestions for future research directions.

II. Background

Topic modeling represents a fundamental technique in unsupervised text analysis, crucial in unveiling concealed thematic structures within a collection of texts. These algorithms are developed to detect patterns of words appearing together and capture the underlying semantic themes that define a group of documents. Topic models identify these themes and allocate them to individual documents. A document encompasses multiple topics, as reflected by its weighted coefficient. The topic with the highest weight dictates the document's primary association, disregarding all other assignments. The subsequent sections discuss the topic models and their assessment using evaluation metrics.

A. *Topic Models*

Topic models produce a list of outputs termed "topic descriptors" using words strongly linked to each topic [12]. From an algorithmic perspective, topics can be envisioned as patterns that emerge from the co-occurrence of words within a given corpus. Numerous algorithms for topic modeling have been developed to address the intricacies inherent in capturing and representing topics within textual data. These algorithms use various mathematical and probabilistic techniques to accomplish their intended objectives.

The development of topic modeling methodologies commenced with Latent Semantic Indexing (LSI), which is referred to as Latent Semantic Analysis (LSA) within the context of topic modeling [13]. LSA employs Singular Value Decomposition (SVD) on the term-document matrix, effectively reducing its dimensionality and capturing latent topics. Although not strictly a probabilistic model, LSA played a pivotal role in the evolution of topic modeling. Another variant of SVD, NMF, was subsequently devised to handle sparse data. NMF dissects the term-document matrix into nonnegative matrices representing topics and their corresponding word distributions [14]. Following this, a probabilistic variant of LSA emerged, known as Probabilistic Latent Semantic Analysis (PLSA), serving as a precursor to the LDA topic model [15]. Across the literature, it has been found that the NMF topic model works better for short text data, whereas LDA is famous for long text data. LDA, the most widely utilized algorithm in topic modeling, has been explored with different sampling methods, and it has been observed that LDA-VB and LDA-CGS are two top-performing variants for topic modeling.

B. *Nonnegative Matrix Factorization (NMF)*

NMF is a non-probabilistic topic model based on the factorization method [16]. In this method, the encoded TF-IDF term-document matrix (sized by M × N) for a given text corpus is decomposed into two matrices: term-topic matrix U and topic-document matrix V, corresponding to K coordinate axes and N points (each point represents one document) in a new semantic space, respectively as shown in Fig. 1.

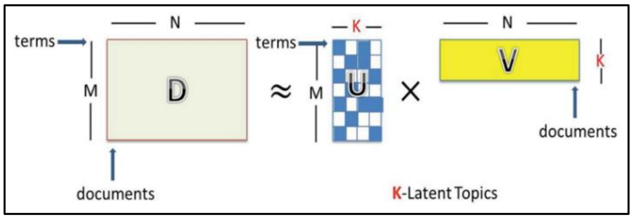


Fig. 1. NMF topic model using factorization method: D ≈ UV, with U and V elementwise nonnegative. [16].

A myriad of diverse applications in various fields use NMF. It is being applied in computational biology to analyze gene expression, unveiling metagenes and their expression patterns [17]. Image processing identifies hidden structures by extracting basis and coding matrices, revealing distinct image components. NMF has been extended to data clustering and pattern discovery across domains through automatic cluster extraction [18].

Concurrently, within the domain of linear algebraic models, there is a consensus among scholars regarding the effectiveness of NMF in handling brief textual content, exemplified by tweets. Notably, NMF's capacity for topic extraction requires no prerequisite knowledge, making it particularly advantageous for research endeavors rooted in social media data [4]. However, being a non-probabilistic topic model, it has limited utility.

C. *Latent Dirichlet Allocation (LDA)*

LDA is a probabilistic topic model postulates that each document consists of a combination of a limited number of topics, each characterized by a word distribution. The primary objective of LDA is to ascertain the optimal topic distribution for each document and the word distribution associated with each topic.

In recent times, the growing demand for topic modeling has stimulated research efforts to enhance the precision and efficiency of inference methods. LDA with Variational Bayes (LDA-VB) assumes a central role within this framework by approximating the posterior distribution governing latent topics and topic proportions. This approximation not only enhances the computational efficiency of LDA but also renders it amenable to the analysis of substantial text datasets [6], in contrast to traditional LDA, which exhibits limitations in computational efficiency and scalability. Despite this distinction, both approaches yield comparable levels of accuracy. Traditional LDA holds an advantage in terms of flexibility, as it accommodates the modeling of hierarchical topic structures, a feature lacking in LDA-VB [19]. LDA's Bayesian framework enhances performance on small or noisy datasets by incorporating prior knowledge, such as known vocabulary, to improve topic identification accuracy. Its flexibility in determining the number of topics makes it suitable for short and long text data, adapting to the data's characteristics. The algorithm for LDA-VB is shown in Fig. 2.

```
Initialize $\lambda^{(0)}$ randomly.
Set the step-size schedule $\rho_t$ appropriately.
repeat
    Sample a document $w_d$ uniformly from the data set.
    Initialize $\gamma_{dk} = 1$, for $k \in \{1, \dots, K\}$.
    repeat
        For $n \in \{1, \dots, N\}$ set
            $\phi_{dn}^{k} \propto exp\{\mathbb{E}[log\, \theta_{dk}] + \mathbb{E}[log\, \beta_{k,w_{dn}}]\}, k \in \{1, \dots, K\}$.
        Set $\gamma_d = \alpha + \sum_n \phi_{dn}$.
    until local parameters $\phi_{dn}$ and $\gamma_d$ converge.
    For $k \in \{1, \dots, K\}$ set intermediate topics
            $\hat{\lambda}_k = \eta + D\sum_{n=1}^{N} \phi_{dn}^{k} w_{dn}$.
    Set $\lambda^{(t)} = (1 - \rho_t)\lambda^{(t-1)} + \rho_t \hat{\lambda}$.
until forever
```

Fig. 2. Algorithm for LDA variational bayes [20].

Among the favored techniques based on sampling for inference, Collapsed Gibbs Sampling (CGS) in conjunction with LDA stands out as a prominent choice. CGS has proven effective and is commonly employed to infer latent topic models [21]. LDA-CGS optimizes LDA for short text data by collapsing latent variables, reducing computational complexity. Empirical evidence suggests LDA-CGS outperforms traditional LDA in various short text applications [22], [23]. The algorithm for collapsed Gibbs sampling is shown in Fig. 3.

```
for $a \leftarrow 1$ to $N$:
        $\boldsymbol{u} \leftarrow$ draw from Uniform $[0,1]$
        for $k \leftarrow 1$ to $K$:
            $P[k] \leftarrow P[k-1] + \frac{\left(N_{kj}^{\neg aj} + \alpha\right)\left(N_{x_{aj}k}^{\neg aj} + \beta\right)}{\left(N_{k}^{\neg ij} + W\beta\right)}$
        for $k \leftarrow 1$ to $K$:
            if $u < P[k]/P[K]$
                then $z_{aj} = k$, stop
```

Fig. 3. Algorithm for Collapsed Gibbs Sampling [24]

LDA finds applications in text classification, document clustering, recommendation systems, topic tracking, and sentiment analysis, enabling categorization, grouping, recommendations, temporal analysis, and sentiment extraction [12]. The following section reviews previous research work for evaluating and comparing topic models.

D. Related Work

Several studies highlight the crucial aspects of assessing topic models using statistical metrics like coherence, stability, diversity, and topic score, comparing them under different conditions using varied datasets.

The comprehensive evaluation performed by Albalawi et al. [7], using the 20-newsgroup dataset and short conversation data from Facebook, sheds light on the efficacy of various Topic Modeling methods. Standard metrics, including recall, precision, F-score, and coherence, were employed to ascertain these methods' ability to generate well-organized and meaningful topics. Notably, this rigorous assessment revealed that two TM methods, LDA and NMF, outperformed the others. Lim et al. [25] explored coherence metrics, introducing a new evaluation approach. They measured differences between topics in different sets of documents and examined how well automated metrics aligned with human judgment, particularly for common topics. Marani et al. [26] focused on enhancing topic model stability by reviewing various methods to measure and improve it. They stressed the importance of considering stability and quality when evaluating topic models. Harrando et al. [27] conducted a comparative analysis of nine popular topic modeling techniques, shedding light on issues in standard evaluation methods [35].

In particular, LDA and NMF exhibited exceptional capabilities in producing diverse and meaningful topic outputs. These findings underscore the prominence of NMF, especially in handling short text data. NMF's superiority suggests its efficacy in addressing the inherent challenges of brevity and noise often found in short texts.

One notable aspect of this research is the focused use of a limited set of standard metrics for evaluation. Rather than delving into an extensive list of metrics, this work highlights the effectiveness of using these fundamental measures to assess topic models.

The suite of experiments in this research work evaluates NMF [14] and variants of LDA in terms of implementation: online Variational Bayes inference (LDA-VB) [28] and inference using Collapsed Gibbs Sampling (LDA-CGS) [16]. These models are selected owing to their popularity and effectiveness across various works in the literature. However, this work aims to identify a topic model that performs effectively and efficiently, irrespective of the text length. The following section focuses on conducting comprehensive experiments to evaluate and compare these prominent topic modeling methods.

III. Experiments

The experimental design provides valuable insights and empirical evidence to draw informed conclusions about the effectiveness and efficiency of chosen topic models. This work uses public datasets to comprehensively compare NMF, LDA-VB, and LDA-CGS-based schemes for short and long text topic mining. All the experiments are conducted on a workstation using Python 3 Google compute engine backend.

A. Datasets

Many datasets exist to assess topic models, showcasing notable variations in corpus size, document length, topic intricacy, and noise levels. The choice of the dataset can profoundly impact the results. Hence, selecting a suitable topic model that aligns with the dataset's attributes is paramount. In the experiment's suite, short and long text datasets are utilized, with additional known associations of labels/topics to assess the performance of model-inferred topics for actual ground truth labels. Short text datasets involve caption datasets, Conceptual Captions, and Wider Captions datasets, comprising captions extracted from the web across various images. The Conceptual Captions [8] dataset is a recently proposed dataset for image captioning comprising instances across multiple categories. Owing to resource constraints, a subset of 10,000 captions with 40 diverse labels were selected. The Wider Captions [9] dataset is another captioning dataset comprising over 50,000 images of events and diverse actions; a sample of 10,000 instances is evaluated. The seminal datasets commonly utilized for topic model evaluation are employed for long text datasets, namely the 20 Newsgroups [10] and the Web of Science [11] datasets. Table I shows that choosing these datasets with varying values of optimal topics and characteristics in terms of document instance sizes and topics allows for a more general evaluation of the models. It can aid in determining the most appropriate model for the task.

TABLE I. DATASETS USED IN EXPERIMENT

Datasets	Documents	Categories/ Topics' K'
Conceptual Captions	10,000/3M	40
Wider Captions	10,000/50,000	61
20-Newsgroups	18,000	20
Web of Science	11697	35

B. *Evaluation Metrics for Assessing Topic Models*

Implementing topic models necessitates making several critical design choices, such as selecting an appropriate algorithm, inference method, model parametrization, and determining the optimal number of topics to uncover. To effectively streamline the process of making these design choices, it becomes imperative to establish a singular, overarching criterion for assessing quality, with accuracy being the foremost consideration. While other factors, such as computational complexity and processing speed, are certainly pertinent, accuracy is paramount to achieving clustering that most faithfully mirrors real-world patterns. Furthermore, the selection of the topic model can vary significantly based on the specific application, and it may yield different outcomes across multiple runs on the same dataset.

Topic modeling evaluation can follow two paths. One involves assessing the internal properties of the clustering result [6] and examining elements like topic-document assignments or topic descriptors corresponding to internal evaluation metrics. These internal metrics scrutinize structural aspects of clusters, such as their separation, without relying on additional input data. However, their quality assessment may not align with human perception, making them the primary choice when a definitive knowledge structure for text clustering is absent. The alternative approach [29] entails comparing clustering results with external knowledge sources, often termed ground truth, which typically takes the form of a pre-defined classification. This classification is often manually assigned and is rooted in human perceptions and the expertise of raters. This approach is known as external evaluation metrics. This research work utilizes internal and external evaluation metrics.

1) Internal evaluation metrics: The internal evaluation metrics can be broadly categorized as Topic Classification, Topic Significance, Topic Coherence, Topic Diversity, and Topic Similarity. Notably, Topic Stability, although an internal measure, is not included in this research as it does not serve as a quality criterion. Instead, it is considered a desirable property for algorithms incorporating stochastic elements [15]. These metrics assess inferred topics' quality, similarity, coherence, divergence, and perplexity. All experiments use the same set of evaluation metrics: internal metrics comprising of diversity [30] and KL-divergence [31] for topic diversity metrics, C_{UMass}, C_V, C_{NPMI}, C_{UCI}, WE-Pairwise and WE-Centroid [30] for topic coherence metrics [32], Jaccard similarity for topic similarity, KL-Uniform, KL-Vacuous, and KL-Background [33] for topic significance, and precision, recall, F1-score, and accuracy [34] for topic-based classification.

Topic Classification Metrics [34] assess document classifier performance using the learned document-topic distribution. This distribution is a K-dimensional representation for training the classifier to predict document classes. Subsequently, the classifier's performance is evaluated using key metrics, including precision, recall, and the F1-Score, which is the harmonic mean of precision, recall, and accuracy. These metrics are calculated using the Formulas in (1) to (4).

$$Precision_i = \frac{c_{ii}}{\sum_{j=1}^{n_c} c_{ji}} \tag{1}$$

$$Recall_i = \frac{c_{ii}}{\sum_{j=1}^{n_c} c_{ij}} \tag{2}$$

$$F1 - Score_i = \frac{2\ X\ Precision_i\ X\ Recall_i}{Precision_i + Recall_i} \tag{3}$$

$$\text{Accuracy} = \frac{\sum_{j=1}^{n_c} c_{ii}}{\sum_{i=1}^{n_c}\sum_{j=1}^{n_c} c_{ij}} \tag{4}$$

where, n_c is the number of classes, and Cij is the confusion matrix.

Topic Significance Metrics [33] play a crucial role in gauging the relevance and importance of topics, with a specific emphasis on scrutinizing both document-topic and topic-word distributions to discern and assess the significance of individual topics [15]. These metrics encompass the following key components:

a) KL-Uniform: This metric compares the topic and W-Uniform distribution using KL-Divergence. The underlying assumption is that genuine topics should be characterized by concisely selecting highly relevant words.

b) KL-Vacuous: In this case, the metric involves evaluating the topic distribution about the W-Vacuous distribution through KL-Divergence. The expectation is that authentic topics should exhibit distinct characteristics compared to a distribution that combines various elements from the sample set.

c) KL-Background: This metric assesses the topic distribution vis-à-vis the W-Background distribution, again utilizing KL-Divergence. The premise here is that genuine topics should only be present in a subset of the documents within the corpus and should not be predominantly composed of background noise.

Topic Coherence Metrics [32] assess how well the top-k words in a topic relate to each other, indicating topic interpretability. This process involves segmentation, probability estimation, confirmation, and aggregation. Standard coherence metrics, detailed in Formulas (5) to (10), are commonly used in the literature. Researchers in past studies have commonly used perplexity or held-out likelihood to evaluate models when comparing different topic numbers 'k.' However, it's crucial to recognize that while perplexity is helpful for this comparison, it primarily assesses predictive performance, not the exploratory goals of topic modeling [16].

$$C_{UCI} = \frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} log\left(\frac{P(w_i,w_j)+\epsilon}{P(w_j)P(w_i)}\right)}{\frac{N(N-1)}{2}} \tag{5}$$

$$C_{UMass} = \frac{\sum_{i=2}^{N}\sum_{j=1}^{i-1} log\left(\frac{P(w_i,w_j)+\epsilon}{P(w_j)}\right)}{\frac{N(N-1)}{2}} \tag{6}$$

$$C_{NPMI} = \frac{\sum_{i=2}^{N}\sum_{j=i+1}^{N} NPMI(w_i,w_j)}{\frac{N(N-1)}{2}} \tag{7}$$

$$\underset{v\ m,\gamma}{\rightarrow}\ (W') = \left\{\sum_{w_i \epsilon W'} m(w_i, w_j)^{\gamma}\right\}_{j=1\ldots|W|} \tag{8}$$

$$C_V = \frac{\sum_{i=1}^{|W|} \vec{v}_{NMPI,1}(W')_i \cdot \vec{v}_{NMPI,1}(W^*)_i}{\left\|\vec{v}_{NMPI,1}(W')\right\|_2 \left\|\vec{v}_{NMPI,1}(W^*)\right\|_2} \tag{9}$$

$$\mathrm{NPMI}(\mathrm{w_i}, \mathrm{w_j}) = \frac{\log\left(\frac{\mathrm{P(w_i,w_j)+\epsilon}}{\mathrm{P(w_j)P(w_i)}}\right)}{-\log(\mathrm{P(w_i,w_j)}+\epsilon)} \tag{10}$$

where,

C_{UCI}	UCI Coherence, based on pointwise mutual information (PMI)
C_{UMass}	UMass Coherence
C_V	newly-proposed coherence measure
NPMI	normalized PMI
W' , W^*	word subsets generated by segmentation
N	number of most probable words per topic
$\mathrm{w_i}, \mathrm{w_j}$	words (specific to a topic)
$\mathrm{P(w_i)P(w_j)}$	word probabilities
$\mathrm{P(w_i, w_j)}$	joint Probability of observing words $\mathrm{w_i}, \mathrm{w_j}$
$\underset{v\ m,\gamma}{\rightarrow}\ (W')$	context vector for words in W', direct confirmation measure m and power γ
ϵ	epsilon for avoiding indeterminate log (0)

In addition to conventional coherence metrics, the research literature has introduced coherence measures for individual topics using word embeddings with the emergence of distributed word representations. These metrics [30], [32] are calculated through pairwise or centroid-based methods, as elaborated in the provided from Formula (11) to Formula (14).

$$W_k = \{w_{ki} \mid w_{ki} \in t_k, i \in argsort\ \{P(w_{kj}) \mid w_{kj} \in t_k\}[: n_{max}]\}\# \tag{11}$$

$$WE - Pairwise_k = \left(\frac{2}{|W_k|(|W_k|-1)} \sum_{i=1}^{|W_k|}\sum_{j=i+1}^{|W_k|} \frac{(e(w_i))^T e(w_j)}{\|e(w_i)\|\|e(w_j)\|}\right);\ w_i, w_j \in W_k\# \tag{12}$$

$$e(w_k^*) = \frac{1}{|W_k|}\sum_{w \in W_k} e(w)\ \# \tag{13}$$

$$WE - Centroid_k = \left(\frac{1}{|W_k|} \sum_{i=1}^{|W_k|} \frac{(e(w_i))^T e(w_k^*)}{\|e(w_i)\|\|e(w_k^*)\|}\right);\ w_i \in W_k\# \tag{14}$$

where,

e(w_i)	Word embedding of word i
n	Maximum number of words per topic in consideration
W_k	Multiset of top n words (in terms of Probability) for topic k
$e(w_k^*)$	Centroid word embedding for the set W_k

Diversity Metrics [30] quantify the variation among the top k-words within a topic, focusing on identifying redundancies by evaluating the recurrence of words. These metrics employ a Symmetric KL-Divergence measure applied to normalized document-topic and topic-word distributions. The primary objective is to assess the diversity within the generated document-topic and topic-word distributions, emphasizing their variability [31]. The mathematical expressions for these diversity metrics, namely KL-Divergence and Topic Diversity, are provided in Formula (15) and Formula (16).

$$\mathrm{KL(R_{l1} \parallel R_{l2})} = \sum_{\mathrm{i=1}}^{\mathrm{T}} \mathrm{R_{l1}} * \log\left(\frac{\mathrm{R_{i1}(i)}}{\mathrm{R_{i2}(i)}}\right) \tag{15}$$

$$\text{Topic Diversity} = \left(\frac{1}{\mathrm{K}}\sum_{\mathrm{k=1}}^{\mathrm{K}} \frac{|\{\mathrm{w_{ki}}|\mathrm{w_{ki}} \in \mathrm{t_k}\ ,\mathrm{i}\in \text{argsort}\ \{\mathrm{P(w_{ki})}|\mathrm{w_{ki}}\in \mathrm{t_k}\}[:\mathrm{n_{max}}]\}|}{\mathrm{n_{max}}}\right) \times 100 \tag{16}$$

where,

K	number of Topics
$\mathrm{t_k}$	topic k
$\mathrm{w_{ki}}$	word i of topic k
$\mathrm{P(w_{ki})}$	probability of word i in topic k as per the topic-word distribution
$\mathrm{n_{max}}$	maximum number of words per topic in consideration

Topic Similarity Metrics [33] come in lexical and semantic forms. Lexical similarity deals with shared word sequences or structures, while semantic similarity relates to shared meaning. Cosine similarity evaluates text similarity by representing documents as term vectors and measuring it as the cosine of the angle between these vectors. Jaccard similarity calculates similarity based on the ratio of shared terms to total unique terms in both texts [3]. Jaccard's similarity [33] is computed to compare topics (Topic A and Topic B) using the Formula in (17).

$$J(Topic\ A, Topic\ B) = \frac{|TopicA \cap TopicB|}{|TopicA \cup TopicB|} \tag{17}$$

where,

$TopicA \cap Topic\ B$	shared words in both topics
$TopicA \cup Topic\ B$	all unique words in both topics

2) External evaluation metrics: On the other hand, external metrics assess topics' performance in terms of classification and clustering of documents based on topic association. The study computed the Adjusted RAND Index (ARI) [29] and the Adjusted Mutual Information (AMI) [35]. While one external clustering metric would typically suffice, both are presented here to compare with findings from other research endeavors. Consider clustering documents as a series of pairwise decisions. If two documents fall in the same class and cluster, or both in distinct classes and clusters, the choice is regarded as accurate; otherwise, it is false. The Rand index calculates the percentage of correct decisions. The adjusted Rand index is the corrected-for-chance version of the Rand index, with an expected value of 0 and a maximum value of 1 for an exact match. On the other hand, the AMI takes a value of 1 when the two partitions are identical and 0 when the MI between two partitions equals the value expected due to chance alone. The mathematical formulation for ARI and AMI is provided in Formula (18) and Formula (19), respectively.

$$\mathrm{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}} \qquad (18)$$

where,

n_{ij},a_i,b_j are values from the contingency table where each entry n_{ij} denotes the number of objects common in clusters.

$$AMI(U,V) = \frac{MI(U,V) - E\{MI(U,V)\}}{max\{H(U),H(V)\} - E\{MI(U,V)\}} \qquad (19)$$

where,

$MI(U,V)$	mutual information between two clusters
$E\{MI(U,V)\}$	expected mutual information between two clusters
$H(U), H(V)$	entropy associated with clusters U and V, respectively

By scrutinizing a range of facets related to the identified topics using internal and external metrics, it becomes feasible to examine the strengths and weaknesses of different topic models in various dimensions or aspects, as discussed in the next section.

C. Experimental Setup

The process of identifying the most suitable topic model adheres to a conventional evaluation approach, encompassing the subsequent stages, as shown in Fig. 4. The topic models selected for consideration were trained separately over each choice of the dataset, followed by an evaluation of the performance of the topics learned by computation of internal and external metrics to determine the most effective topic model across datasets with diverse characteristics. Each data set is preprocessed to eliminate irregular word forms across documents. The topic models were trained over the preprocessed datasets with the best possible choices of hyperparameters: the number of topics being the optimum for the dataset, along with other parameters selected appropriately (Dirichlet priors selected to be symmetric and as per the value of number of topics). Following the training of topic models, the models were evaluated using internal and external metrics covering various aspects of topic quality: coherence, significance, diversity, similarity, usability for classification, and clustering information. The process of training and evaluation was jointly performed by use of the OCTIS framework [5], which incorporates implementations of multiple topic models and provides a suite of diverse evaluation metrics for comparing and contrasting the same. The experiment was repeated for all datasets chosen for this work. Within each experiment, apart from listed internal metrics covering topic diversity, topic coherence, topic significance, topic classification, and topic similarity, external metrics covering the extent of information conveyed by topics for determining a suitable clustering were also employed to determine the efficacy of learned topics.

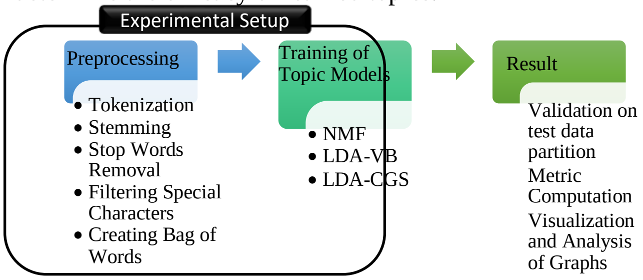


Fig. 4. Stages of experimental setup and result.

1) Data preprocessing: Each examined dataset has undergone preprocessing to eliminate irregular word forms across documents. This preprocessing encompasses standard procedures such as tokenization, stemming, removing stop words, and filtering special characters. Additionally, the preprocessing involves generating a bag-of-words representation that can be utilized in various topic models.

2) Training of topic models: Following the essential preprocessing phase, the dataset is divided into distinct training and testing segments, each allocated with specific roles and objectives. The training portion is the foundation for training the selected topic models, enabling them to gain insights and patterns from the provided data. On the other hand, the testing portion plays a pivotal role in the evaluation process, serving as an independent set of documents against which the models' performance is scrutinized and assessed.

During the evaluation phase, each chosen topic model is subjected to rigorous scrutiny over these training and testing datasets. This assessment entails meticulously examining their output in light of internal and external evaluation metrics. Internal evaluation metrics encompass criteria that gauge the coherence, diversity, and overall quality of the topics generated by the models. Topic classification metrics are another category of popular metrics employed to assess the quality of topic models by assessing the extent of how useful topics are as features for the classification of the source documents using standard classification techniques. Our experiments observed poor performance of the learned topics for classifying documents using an SVM classifier. Thus, we have included only a relative comparison based on the values highlighting the best topic model per dataset and metric. A possible explanation for the poor metric values is the sensitivity of the metrics to the choice of classifier and the corresponding hyperparameters. So, these metrics may not accurately represent the usefulness of topics for classifying documents. A thorough analysis for contrasting performance must also compare diverse classifiers to determine the overall effectiveness of topics learned by the different topic models for classification, which is beyond the scope of this study of contrasting topic models.

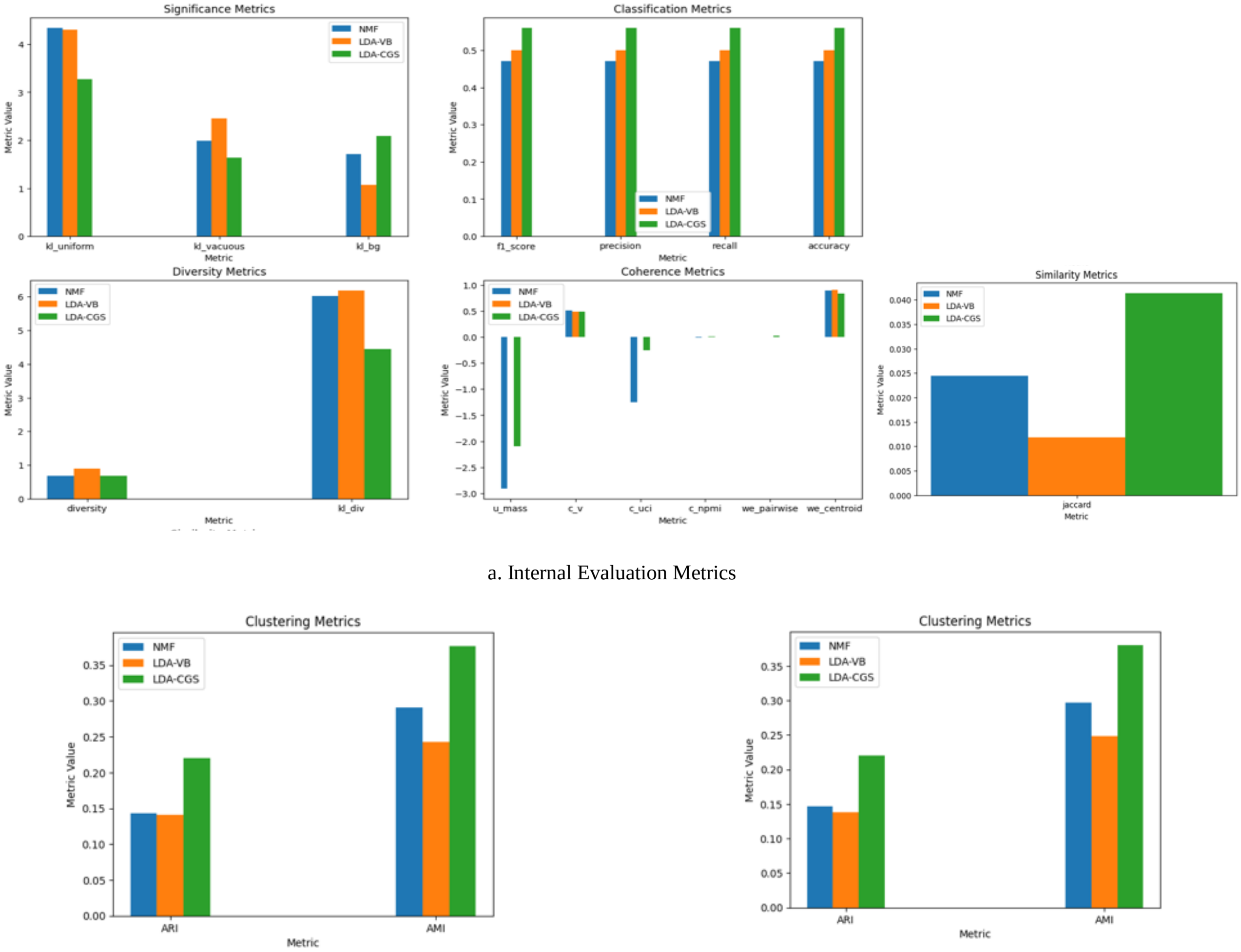
External evaluation metrics, on the other hand, pivot towards assessing the model's efficacy in practical applications. They evaluate the model's ability to categorize, classify, and cluster documents based on the topics it infers. This dual-pronged evaluation strategy, involving internal and external metrics, forms a comprehensive and well-rounded approach to ascertain the effectiveness and applicability of the topic models under scrutiny.

Once the models have been trained successfully, they are run on test data. The output is analyzed using the graphs discussed in the next section.

IV. Results and Discussion

A. Results over Short Text Datasets

Varying results regarding the most performant topic model across different internal metrics are observed across short text datasets. Regarding internal metrics, NMF is significantly more performant than LDA implementations across many metrics. Across external metrics, however, LDA-CGS outperforms both NMF and LDA-VB for both datasets.



a. Internal Evaluation Metrics

b. External Evaluation Metrics – Training Partition

c. External Evaluation Metrics – Test Partition

Fig. 5. Comparison of topic models over Internal and External Metrics for the Conceptual Captions dataset

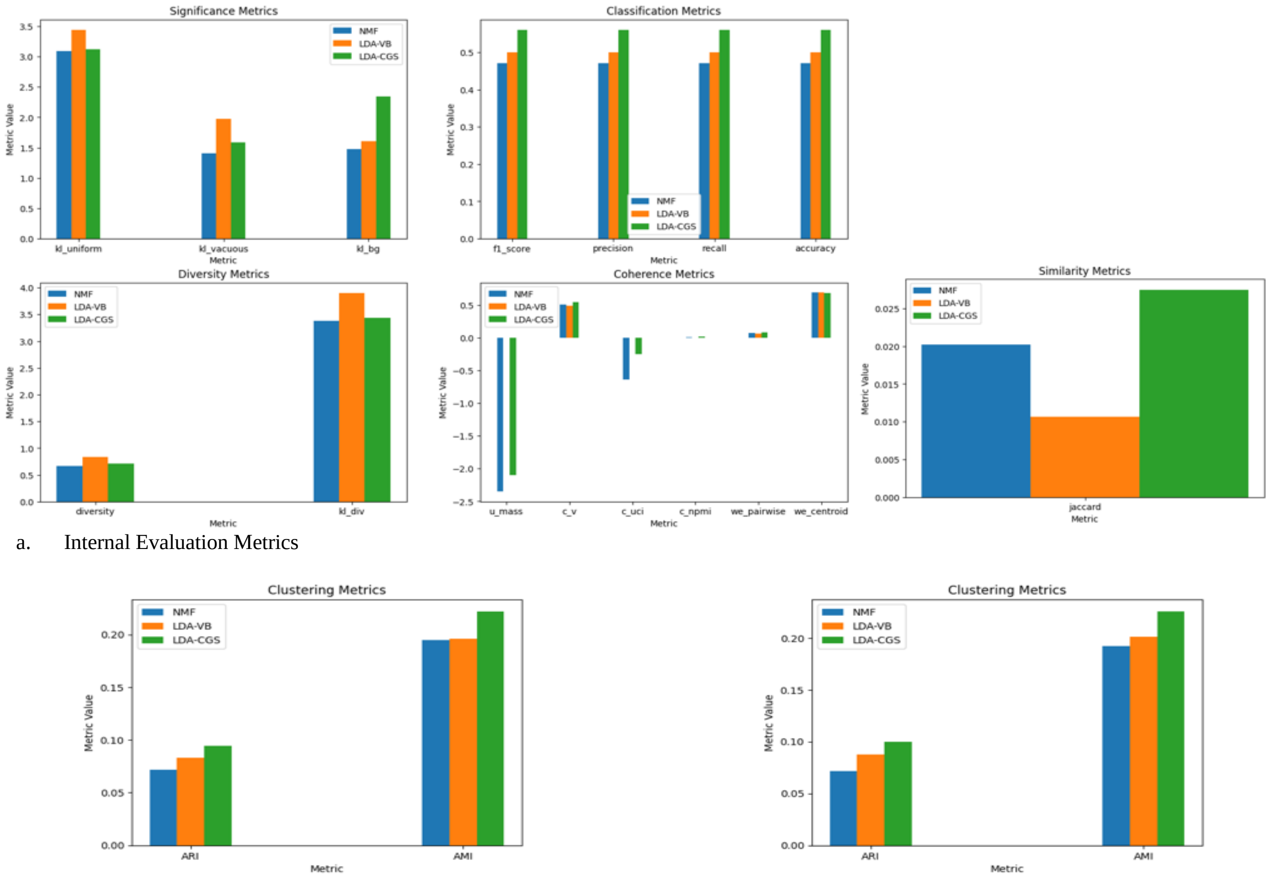
Fig. 5(a), 5(b), and 5(c) compare the performance of the three topic models over the Conceptual Captions dataset regarding internal and external evaluation metrics for training and test partitions, respectively. Across internal metrics, while there exists no consensus for the best topic models, NMF demonstrates high performance across a significant number of metrics. This behavior is prominent in topic classification metrics, where NMF outperforms both LDA variants by a minute margin. NMF also demonstrates leading performance across coherence metrics C_V and C_{NPMI} and topic significance compared to the Uniform distribution (KL-Uniform). Across other metrics, while NMF is not the best model, the difference between the metric values of NMF and that of the best model is minimal. Across the remaining metrics, LDA-CGS and LDA-

VB perform in contrast to one another across different metrics. While the topics of LDA-CGS are more coherent in terms of C_{UMass} and WE-Pairwise, more significant in terms of KL-Divergence for the Background distribution, the topics are more similar and less diverse than LDA-VB topics. Across external evaluation metrics, however, LDA-CGS is the best performant, demonstrating an ability to produce a prominent quality clustering based on inferred topics, with an agreement to the ground truth of the document labels.

The performance of the topic models in terms of internal and external metrics for the Wider Captions dataset over training and test partitions is compared in Fig. 6(a), 6(b), and 6(c), respectively. For the dataset, the behavior of NMF is much more prominent, with leading results across many metric categories. NMF outperforms both variants of LDA across topic significance (KL-Uniform), topic classification, topic similarity, and topic diversity when measured with KL-divergence. However, unlike prior observations, topics inferred by NMF for the dataset are not as coherent as those inferred by other models, and instead of NMF, LDA-CGS results in the most coherent topics as observed by the coherence values for all coherence metrics. However, the divergence and similarity of the LDA-CGS topics are still significantly lower than the other models. LDA-VB demonstrates comparable results to the most compelling topic model across most metrics and leads over other models only in terms of diversity and topic significance (KL-Vacuous). Across external metrics, a similar trend as the Conceptual Captions dataset is observed, with LDA-CGS significantly outperforming other topic models, demonstrating the reasonable effectiveness of LDA-CGS in inferring high-quality topics that produce high-quality clusters.

B. Results over Long Text Datasets

Across long text datasets, NMF is superseded by LDA models across internal and external evaluation metrics. It demonstrates the effectiveness of LDA in general; most corpora on which topic modeling is utilized comprise long text documents with multiple topics per document. LDA is significantly more effective than NMF, and short text document corpora are rare and application-specific.

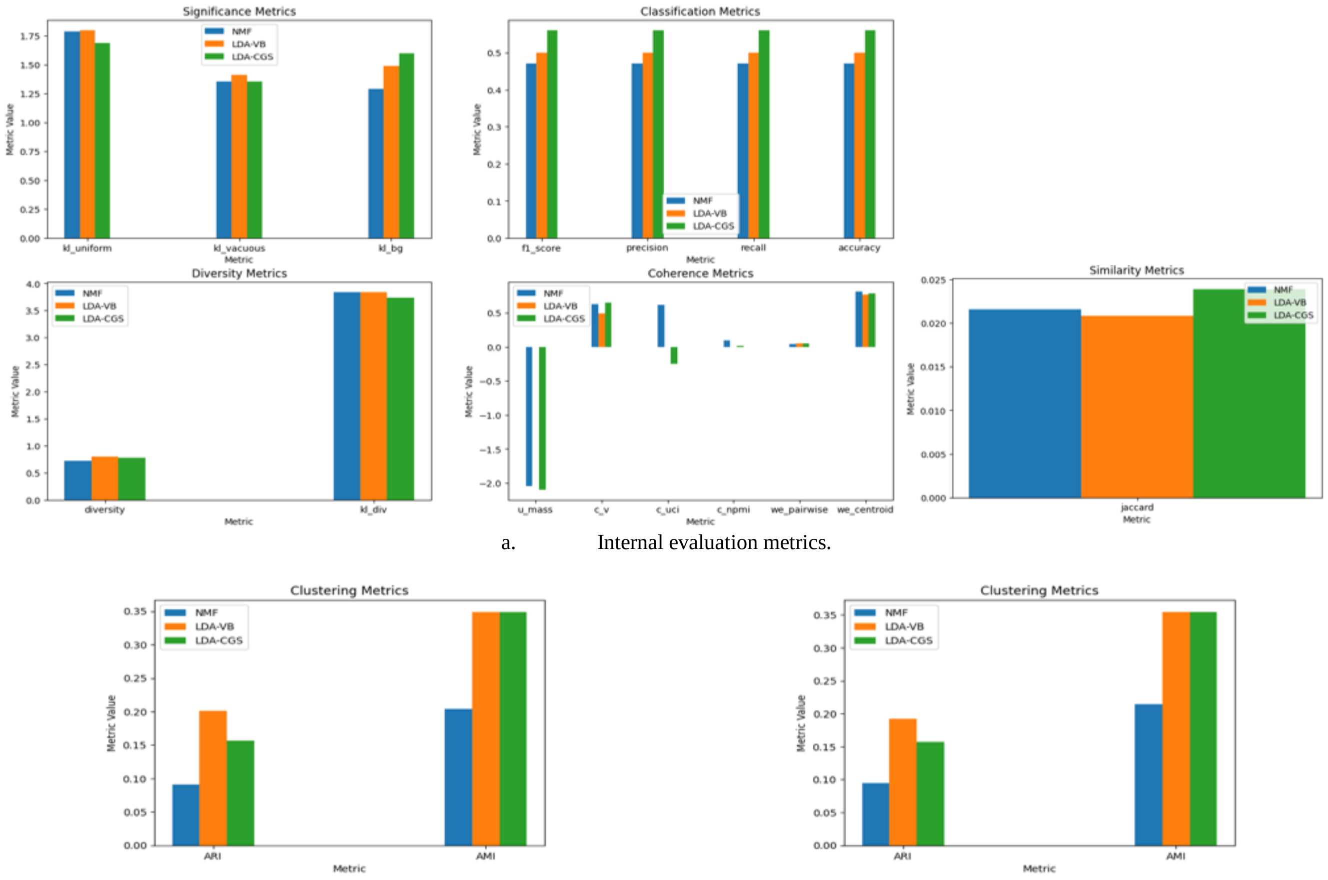


a. Internal Evaluation Metrics

b. External Evaluation Metrics – Training Partition

c. External Evaluation Metrics – Test Partition

Fig. 6. Comparison of topic models over internal and external metrics for the wider captions dataset.



a. Internal evaluation metrics.

b. External evaluation metrics – training partition.

c. External evaluation metrics – test partition.

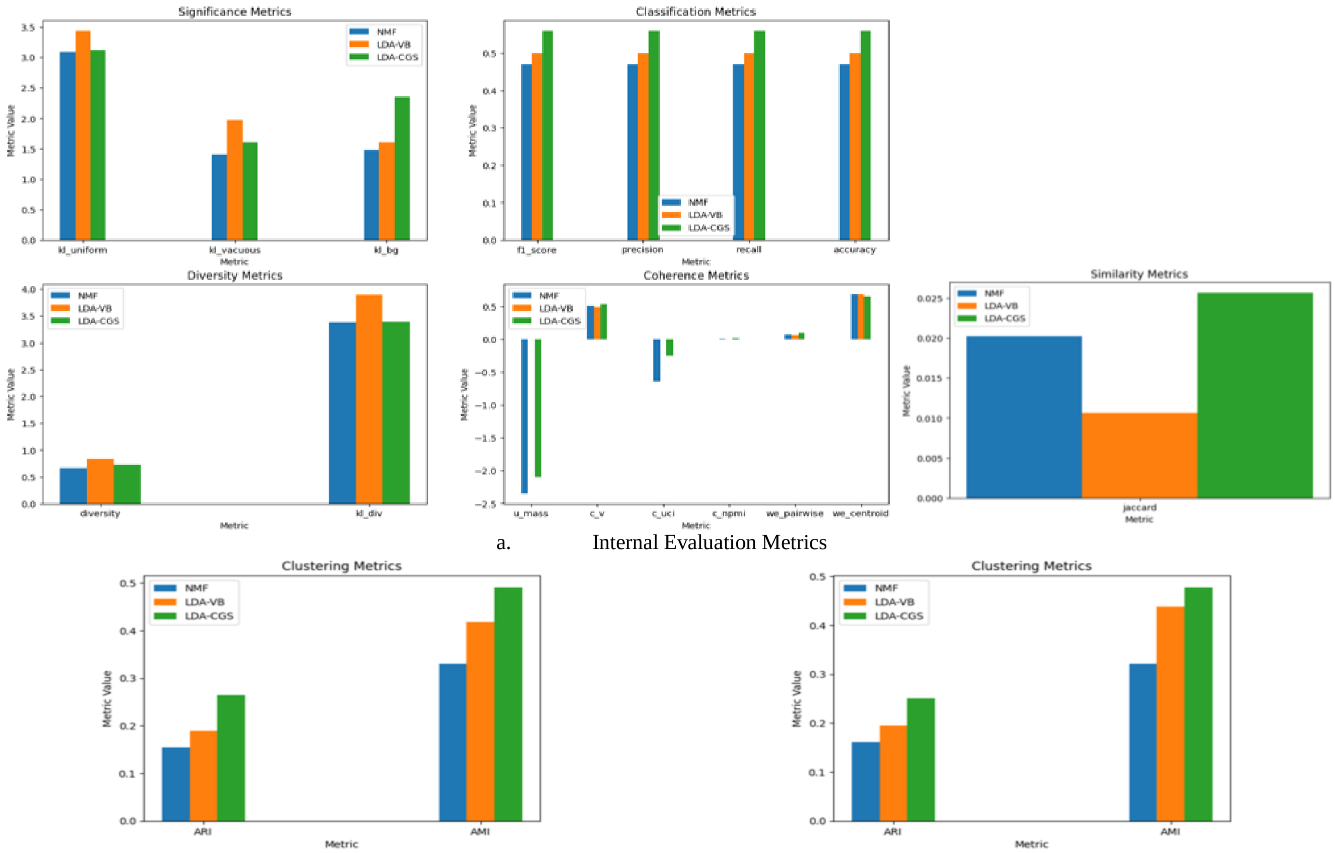
Fig. 7. Comparison of topic models over internal and external metrics for the 20-newsgroups dataset.

Fig. 7(a), 7(b), and 7(c) compare the performance of the three topic models over the 20-Newsgroups dataset regarding internal and external evaluation metrics over training and test partitions, respectively. Regarding internal evaluation metrics, the results regarding determining the best topic model are inconclusive as the metrics appear to disagree. However, across all metrics, NMF is outperformed by either variant of LDA. Classification and Diversity metrics prefer topics inferred by the variational Bayes implementation of LDA. In contrast, topics produced by LDA implemented with Collapsed Gibbs Sampling are found to be the most coherent and significant compared to the background distribution. Across external evaluation metrics, LDA-CGS outperforms LDA-VB in terms of AMI but not in terms of ARI scores. Both variants of LDA perform similarly for this dataset, and either can be preferred for inferring latent topics. It is noted that the difference across metric values for both LDA variants is insignificant, strengthening the assertion that either model is suitable for obtaining high-quality topics.

The comparison of the performance of topic models over the Web of Science dataset in terms of internal and external metrics over training and test partitions is demonstrated in Fig. 8(a), 8(b), and 8(c) respectively. Similar to the trend across other datasets, internal evaluation metrics do not demonstrate a unified consensus for the choice of the topic model. However, LDA variants significantly outperform NMF in all metrics. However, unlike 20-Newsgroups, LDA-CGS is found to be more performant than LDA-VB across classification and coherence metrics. The trend for significance and diversity metrics, however, is similar to the trend observed in 20-Newsgroups and other short text datasets, which indicates that LDA-CGS, in general, infers topics that are highly coherent and more significant than background information but are not significantly diverse in general. Further, the trend across external metrics is similar to the observations across short text datasets, with LDA-CGS outperforming other topic models, demonstrating the reasonable quality of LDA-CGS inferred topics and their effectiveness in clustering documents.

C. *Discussion and Interpretation of Results*

Based on the results, LDA-CGS is suitable for producing high-quality topics across corpora with diverse text characteristics. However, to improve the quality of topics specific to the application, LDA-CGS produces reasonable quality topics that demonstrate high performance over external clustering metrics and classification over documents, especially with long text data. Further, the topics have high coherence, divergence, significance, and reasonable similarity, making the algorithm a suitable default choice for most applications. The best performant topic model across each dataset and metric has been summarized in Table II.



a. Internal Evaluation Metrics

b. External evaluation metrics – training partition.

c. External evaluation metrics – test partition.

Fig. 8. Comparison of topic models over internal and external metrics for the web of science dataset.

TABLE II. COMPARISON OF TOPIC MODELS NMF, LDA-VB, AND LDA-CGS) ACROSS INTERNAL AND EXTERNAL EVALUATION METRICS SET

Evaluation Metrics		**Datasets**			
Metric Category	**Metric**	**Conceptual Captions**	**Wider Captions**	**20-Newsgroups**	**Web of Science (WOS11697)**
Topic Classification [9] [10]	**Precision**	NMF	NMF	LDA-VB	LDA-CGS
	Recall	NMF	NMF	LDA-VB	LDA-CGS
	F1-Score	NMF	NMF	LDA-VB	LDA-CGS
	Accuracy	NMF	NMF	LDA-VB	LDA-CGS
Topic Significance [9] [13]	**KL-Uniform**	NMF	NMF	LDA-VB	LDA-VB
	KL-Vacuous	LDA-VB	LDA-VB	LDA-VB	LDA-VB
	KL-Background	LDA-CGS	LDA-CGS	LDA-CGS	LDA-CGS
Topic Coherence [14] [20]	**C_U_Mass**	LDA-CGS	LDA-CGS	LDA-CGS	LDA-CGS
	C_UCI	LDA-VB	LDA-CGS	LDA-CGS	LDA-CGS
	C_NPMI	NMF	LDA-CGS	LDA-CGS	LDA-CGS
	C_V	NMF	LDA-CGS	LDA-CGS	LDA-CGS
	WE-Pairwise	LDA-CGS	LDA-CGS/ LDA-VB	LDA-CGS	LDA-CGS
	WE-Centroid	LDA-VB	NMF	NMF	LDA-VB/ NMF
Topic Diversity [21] [29]	**KL-Divergence**	LDA-VB	NMF	LDA-VB	LDA-VB
	Diversity	LDA-VB	LDA-VB	LDA-VB	LDA-VB
Topic Similarity [33]	**Jaccard Score**	LDA-VB	NMF	LDA-VB	LDA-VB
External Topic / Clustering Quality [12] [32]	**Adjusted Rand Index (ARI)**	LDA-CGS	LDA-CGS	LDA-VB	LDA-CGS
	Adjusted Mutual Information (AMI)	LDA-CGS	LDA-CGS	LDA-CGS	LDA-CGS

V. Conclusion

Topic modeling techniques exhibit versatility in handling both short and long text data. While models like LDA excel with longer documents, approaches such as NMF demonstrate efficiency in understanding the context within shorter texts. This research performs empirical analysis of state-of-the-art topic models through statistical metrics in the context of varied text length data to determine the best model irrespective of text length. Based on the experimental results, LDA-CGS produces high quality topics over external clustering metrics for both long and short text data. The topics produced by LDA-CGS have high coherence, divergence, significance, and similarity, making it a suitable choice for most datasets.

Future research directions could explore hybrid approaches that combine the strengths of multiple algorithms to enhance topic modeling performance across different text lengths. The results show that NMF is still a strong contender for short text data, and a hybrid model may show much better performance for varying text length datasets. These models can also be applied on Twitter or image caption datasets to discover relevant information for the classification process.

Conflict of Interest: The authors declare no competing interests and confirm that neither the manuscript nor any parts of its content are currently under consideration or published in another journal.

References

[1] B. A. H. Murshed, S. Mallappa, J. Abawajy, M. A. N. Saif, H. D. E. Al-ariki and H. M. Abdulwahab, Short-text topic modelling approaches in the context of big data: taxonomy, survey, and analysis, vol. 56, Springer Science and Business Media LLC, 2022, p. 5133–5260.

[2] S. Sbalchiero and M. Eder, Topic modeling, long texts and the best number of topics. Some Problems and solutions, vol. 54, Springer Science and Business Media LLC, 2020, p. 1095–1108.

[3] S. Athukorala and W. Mohotti, An effective short-text topic modelling with neighbourhood assistance-driven NMF in Twitter, vol. 12, Springer Science and Business Media LLC, 2022.

[4] R. Egger and J. Yu, A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts, vol. 7, Frontiers Media SA, 2022.

[5] S. Terragni, E. Fersini, B. G. Galuzzi, P. Tropeano and A. Candelieri, OCTIS: Comparing and Optimizing Topic models is Simple!, Association for Computational Linguistics, 2021.

[6] M. Rüdiger, D. Antons, A. M. Joshi and T.-O. Salge, Topic modeling revisited: New evidence on algorithm performance and quality metrics, vol. 17, D. R. Amancio, Ed., Public Library of Science (PLoS), 2022, p. e0266325.

[7] R. Albalawi, T. H. Yeap and M. Benyoucef, Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis, vol. 3, Frontiers Media SA, 2020.

[8] P. Sharma, N. Ding, S. Goodman and R. Soricut, Conceptual Captions: A Cleaned, Hypernymed, Image Alt-text Dataset For Automatic Image Captioning, Association for Computational Linguistics, 2018.

[9] Y. Xiong, K. Zhu, D. Lin and X. Tang, Recognize complex events from static images by fusing deep channels, IEEE, 2015.

[10] K. Lang, NewsWeeder: Learning to Filter Netnews, Elsevier, 1995, p. 331–339.

[11] K. Kowsari, Web of Science Dataset, Mendeley, 2018.

[12] A. Goyal and I. Kashyap, Latent Dirichlet Allocation - An approach for topic discovery, IEEE, 2022.

[13] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer and R. Harshman, Indexing by latent semantic analysis, vol. 41, Wiley, 1990, p. 391–407.

[14] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," Nature, vol. 401, p. 788–791, October 1999.

[15] D. Rugeles, Z. Hai, J. F. Carmona, M. Dash and G. Cong, Improving the Inference of Topic Models via Infinite Latent State Replications, arXiv, 2023.

[16] Y. Chen, H. Zhang, R. Liu, Z. Ye and J. Lin, Experimental explorations on short-text topic mining between LDA and NMF based Schemes, vol. 163, Elsevier BV, 2019, p. 1–13.

[17] K. Devarajan, Nonnegative Matrix Factorization: An Analytical and Interpretive Tool in Computational Biology, vol. 4, B. Bryant, Ed., Public Library of Science (PLoS), 2008, p. e1000029.

[18] Z.-Y. Zhang, Nonnegative Matrix Factorization: Models, Algorithms and Applications, Springer Berlin Heidelberg, 2012, p. 99–134.

[19] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li and L. Zhao, Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey, vol. 78, Springer Science and Business Media LLC, 2018, p. 15169–15211.

[20] M. D. Hoffman, D. M. Blei, C. Wang and J. Paisley, Stochastic Variational Inference, vol. 14, 2013, p. 1303–1347.

[21] A. U. Rehman, Z. Rehman, J. Akram, W. Ali, M. A. Shah and M. Salman, Statistical Topic Modeling for Urdu Text Articles, IEEE, 2018.

[22] Y. W. Teh, D. Newman and M. Welling, "A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation," in Advances in Neural Information Processing Systems 19, The MIT Press, 2007, p. 1353–1360.

[23] M. O. Ajinaja, A. O. Adetunmbi, C. C. Ugwu and O. S. Popoola, "Semantic similarity measure for topic modeling using latent Dirichlet allocation and collapsed Gibbs sampling," Iran Journal of Computer Science, vol. 6, p. 81–94, November 2022.

[24] I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth and M. Welling, Fast collapsed gibbs sampling for latent dirichlet allocation, ACM, 2008.

[25] J. P. Lim and H. Lauw, "Large-Scale Correlation Analysis of Automated Metrics for Topic Models," in Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023.

[26] A. Hosseiny Marani and E. P. S. Baumer, "A Review of Stability in Topic Modeling: Metrics for Assessing and Techniques for Improving Stability," ACM Computing Surveys, vol. 56, p. 1–32, November 2023.

[27] I. Harrando, P. Lisena and R. Troncy, "Apples to Apples: A Systematic Evaluation of Topic Models," in Proceedings of the Conference Recent Advances in Natural Language Processing - Deep Learning for Natural Language Processing Methods and Applications, 2021.

[28] M. Hoffman, F. Bach and D. Blei, Online learning for latent dirichlet allocation, vol. 23, 2010.

[29] J. Han, M. Kamber and J. Pei, Data mining concepts and techniques third edition, 2012.

[30] A. B. Dieng, F. J. R. Ruiz and D. M. Blei, Topic Modeling in Embedding Spaces, arXiv, 2019.

[31] F. Bianchi, S. Terragni and D. Hovy, Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence, arXiv, 2020.

[32] M. Röder, A. Both and A. Hinneburg, Exploring the Space of Topic Coherence Measures, ACM, 2015.

[33] L. AlSumait, D. Barbará, J. Gentle and C. Domeniconi, Topic Significance Ranking of LDA Generative Models, Springer Berlin Heidelberg, 2009, p. 67–82.

[34] X.-H. Phan, L.-M. Nguyen and S. Horiguchi, "Learning to classify short and sparse text {\&}amp$\mathsemicolon$ web with hidden topics from large-scale data collections," in Proceedings of the 17th international conference on World Wide Web, 2008.

[35] S. Romano, N. X. Vinh, J. Bailey and K. Verspoor, Adjusting for Chance Clustering Comparison Measures, arXiv, 2015.

Assessing User Requirements for e-Resources Interface Design in University Libraries

Yuli Rohmiyati[1], Tengku Siti Meriam Tengku Wook[2], Noraidah Sahari[3], Siti Aishah Hanawi[4]
Faculty of Information Science and Technology, University Kebangsaan Malaysia, Selangor, Malaysia[1, 2, 3, 4]
Department of Library Science, Diponegoro University, Semarang, Indonesia[1]

***Abstract*—e-Resources in the university library as learning resources are one of the primary services that promote learning and research to improve university productivity. At present, users find it difficult to access e-resources and require assistance in finding them. When using the system, users felt frustrated, confused, and lost. The e-resources services system on library websites, on the other hand, lacks sociability and a sense of human warmth. Sociability and a sense of human warmth can be integrated into the website interface, which may evoke the sensation of being with an actual individual, even if the service is provided online. This study investigated the social presence aspects that can be implemented in the library's e-resources system. The purpose of this study is to elicit social presence features that can be implemented in the design of e-resource interfaces on library websites. The methods used in this study are in three phases: a) web content analysis from twelve university library interfaces designed in several countries; b) interviews with library staff; and c) assessment by a questionnaire of library website users. Website content analysis was used to investigate elements that offer many unique features to support the implementation of social presence through the e-resources interface. An interview was used to validate elements that were found in the web content analysis, and a questionnaire phase was used to assess the user requirements for these social presence elements. The results of empirical studies show that users need some elements of social presence, such as comments, chat, ratings, voice, personalized welcome in library accounts, tools, preference language, links for reference managers, and social media, as well as ease of access such as readable help font, color, and font size.**

Keywords*—User requirement; interface design; e-resources; social presence; university library; element

I. Introduction

e-Resources in the library websites prefer to provide authentic learning with social interaction [1], and a user interface is a communication tool for the user with the e-resources system in the library websites. As one of the official sources of information, which is one of the services on the library website, are often requested to meet a user's needs quickly, which was found to be frustrating [2], boring [3], confusing [4], and feeling lost [5]. It is because the design interface of e-resources has unclear navigation [6] with the searching interface of e-resources [7] and with too much information [8]. This makes it hard for users to find information [9] [10]. Users feel difficulty in accessing the system and understanding the system navigation [5], added to a lack of a guide and the absence of some websites [11]. The other reason that causes frustration for the user is a slow internet network [12] [13] [14], low bandwidth [15], slow download [16], and limited power [17]. All these reasons make a lack of time [18] with high access costs [19], lack of authority, and also limited subscription titles [20], which makes the low use [21] of e-resource systems in libraries.

However, the provision of user interface designs in e-resources services needs to be addressed more [22], to sustain the use of e-resources especially in the university library systems [23] [3]. Among them is navigation on the interface that needs to be clarified and made easier to identify and access complex systems. Meanwhile, the e-resources in the university library is the primary source of information obtained by students and the academic community. [24] [25] believes that e-resources are the main gateway that provides users with the information needed to conduct investigations in various fields. University e-resources are the main part of the library websites and have become a vital academic resource that supports teaching, learning, and research activities [26]. e-Resources play an essential role as a source of information in providing quick and easy service to readers [13]. In contrast, librarians [27] play a role in facilitating the discovery and access to information resources.

The challenge here is, firstly, that e-resources interfaces lack elements of social activity or human-friendly treatment from face-to-face or online service experience [28] [29], which causes users to feel confused, and frustrated and to resolve the problems encountered themselves [30]. Studies by [31] and [32] found that 87.2 % of users expect multiple database searches to be available on one interface. Most users prefer or tend to use open-access journals as opposed to journals that need to be subscribed to [33] [34]. In addition, when users do not get answers to their questions and there is no help in the system, users will solve their difficulties while using the system and spend a long time solving the problems encountered. Users feel depressed and frustrated because users do not get treated with face-to-face services.

Secondly, the system is less friendly. The cognitive and affective issues faced by these users are due to the lack of information about the social elements on the system interface [35] [36]. This indicates that the current system does not emphasize or integrate elements that evoke a sense of social presence through their system interfaces. Previous studies suggest social cues such as voice, picture, or video [37], and greetings influence cognitive and affective elements [38] [36]. The use of social elements in a system is referred to as social presence [39]. Using the elements or characteristics of social activities can reduce users' negative experiences using the system [40]. Therefore, a sociable and human warmth user interface is needed to make users engage with the e-resource system in the library.

Thirdly, e-resources are all digital collections that can be the main reference for students and lecturers to improve knowledge, education quality, and university performance. Based on the results of previous studies, it is known that the use of these e-resources is low [41] [42] [43]. The cause is the absence of sociability and human warmth in this interface system. Because of that, users turn to ordinary search engines because they feel comfortable with them compared to using e-resources in libraries. Users prefer to do searches or get information on standard search engines [20]. Although library initiatives provide other mobile application libraries [44], the usage is still at a low stage [24]. Meanwhile, the system's performance makes it easier [45] for anyone to access information and receive service.

Fourthly, social and technological factors help increase the use of learning systems [46]. Social activities in this system include greetings [28] [47], chats [48] [49], and images [50] [51]. According to [52], presenting elements of social activity in the system will influence users to engage in the cognitive and effective use of e-resources. Although studies on the use of e-resources in libraries have been carried out by many researchers [2] [53] studies that focus on social presence in e-resources interface design are limited. Therefore, analyzing social presence elements on the interface of the system or pages in the library is a critical issue to be investigated to see the benefits that can be provided in the environment. Previous studies show that social presence was applied to e-learning, e-commerce, and e-services. The concept of social presence can even connect users with technology in a fun way.

Based on the above, this study aims to investigate the element of social presence to be applied and incorporated into the e-resources interface design in the university libraries. Investigation of this element is needed to suit users' needs for e-resources which are undoubtedly different from users' needs in e-commerce or e-learning. The research question in this study are:

1) What elements are needed for the university library's e-resources interface design?

2) How to assess these elements to meet the user requirement for e-resources interface design in the university library?

To achieve this, this study has designed a three-phased method approach. Firstly, this study finds elements from the literature reviews and searches for some elements from web content analysis. Secondly, this study validates all elements from the first phase by trying the element's function and also conducting an interview session with some librarians from these libraries. Thirdly, this study assesses that element to find user requirements through a survey. This investigation can develop new elements needed in designing an e-resource interface in university libraries by completing this phase.

The results of this research reveal the user requirements of e-resources service interface design in university libraries. The scope and contribution of this research consists of two main areas, namely human-computer interaction and libraries. The structure of this article is as follows: an introduction in Section I to the issues of e-resource usage, followed by an explanation of related work in Section II, methodology in Section III and the result of the user requirement in Section IV. Finally discussion and conclusion is mentioned in Section V and Section VI respectively.

II. Literature Review

This study explores research in the areas of accessibility of e-resources and social presence. This study summarizes the literature and presents it in a literature review.

A. Accessibility of e-Resources

Accessibility of e-resources refers to the use of e-resources subscribed by the library by all users to access the information and services provided through the website [45]. This accessibility describes the extent to which resources, services, and products are accessible to the users [54] who utilize them or the ability to access information stored by the operator by minimizing any distance and cost barriers by using a user interface. Accessibility determines the speed of the output of information that can be accessed in any format [55]. Access to information is essential in determining successful and practical research in universities. Factors that influence the access and use of electronic information resources in university libraries are:

1) Literacy in obtaining information is of paramount importance in quality research. Literacy helps users access and use relevant information. Most users need these skills, which causes them to spend more time retrieving the necessary information. These skills include knowledge of database structure, appropriate and available search terms, and an understanding of how the commands are linked to each other.

2) Consumer attitude towards electronic information sources. A previous study in [3] found that lecturers prefer electronic information sources to printed ones. Electronic information sources require less time to access such information sources than printed information materials.

3) Access and use of electronic information resources. Research results show that electronic information resources are still low even if accessed using library computers or remote access [56]. It is common knowledge that there are various benefits of using e-resources, namely: (1) Access to a broader collection of information, (2) Faster access to information, (3) Increased Academic Performance, (4) Access to quality information, (5) Access to the latest information and (6) Easier access to information.

The challenge of using e-resources in libraries is that it takes much time to find the materials needed [57] and it is not easy to access. This finding can be attributed to inaccessible web design [58], low level of literacy in using e-resources, lack of knowledge in using assistive technology, and lack of motivation, which may be due to the above factors.

B. Social Presence

Social presence is a concept that has its roots in telecommunication literature. The study of Short, Williams, and Christi developed the social presence theory as a model to analyze socio-psychological dimensions of mediated communication from the perspective of social cues [50] [48]. They defined social presence as the degree of excellence of others in interactions and the importance of impressions of interpersonal relationships.

Social presence is a person's mental state to feel close to others in a virtual environment [59] [60]. At the same time, social presence can also be defined as a person's participation as a real person in expressing self and social emotions through communicative media [61] [62]. A person can tap into a high level of social presence, escape from the real world, and enter an exhilarating and pleasurable mental state through intimate interactions with other users. With a high social presence, a person tends to feel more satisfied and comfortable using the system. Someone with a strong social presence in the social networking gets a more extraordinary experience. A strong sense of excitement helps users enter an exhilarating mental state when using something they enjoy. In another study, belongingness [60] was positively related to well-being embodied by happiness and pleasure.

Short stated that social presence contributes to the medium of communication. Social presence has a complex structure, like the participation of someone as a real person in expressing himself and their emotions through communicative media [62]. Social presence is a construct consisting of two concepts: intimacy and immediacy. Intimacy is a sense of closeness one builds to another through verbal and non-verbal cues. Whereas; immediacy is the frequency of reactions and interactions of a person willing to support and share experiences with others. [63] [30] state that if someone in a system demonstrates these two concepts it can increase social presence.

Social presence is a theory that has its foundation in telecommunications literature. Short, Williams, and Christi developed social presence theory as a model for analyzing the socio-psychological dimensions of mediated communication from the perspective of emerging social activities [50] [63]. They define social presence as a person's level of superiority when interacting and the impact of those interpersonal relationships. Moreover, according to [64], intimacy in communication media is influenced by physical distance, eye contact, smiles, and personal topics of conversation.

The following are the elements of social presence obtained from a literature review (see Table I).

TABLE I. SOCIAL PRESENCE ELEMENT FROM A LITERATURE REVIEW

Author	Social Presence elements
S. S. Engku Alwi and T. S. M. Tengku Wook	1) Services: Communication with retailers, Feedback/ Comments, Frequently Asked Questions, Chat. 2) Display: Language choices, Advertising. 3) Website content: Photos, Animation, Audio, Video, Automotive description, Product description. 4) Additional applications: Recommender system, Rating, User review [65]
Thabet dan Zghal	Website data, human photos, personal responses, animations, virtual agents, frequently asked questions (FAQ), instant messaging, entries, discussion forums, and social networks.
Papadopoulou and Ganguly et al.	Images, colors, and graphics
Y. Li and Y. Xie	Visuals (images and colors) [66]
Y. J. Kang and W. J. Lee	Avatar [67], animated gif, username, greeting [28]
W. Nadeem et al.	Comments, ratings, exchange options, and tagging [68].
J. Wei et al.	Communication support (i.e., text chat and voice chat) and shared navigation (i.e., page push) [60]

Based on Table I, it can be seen that the social presence elements that can be applied in designing the electronic resource interface on the university library website include: communication with library staff, chat, language choice, photo, and greeting, welcoming users by name, text chat, and emoticons.

III. METHODS

This research method through three phases (see Fig. 1). The primary purpose of this article is to identify social presence elements for the e-resources interface in the university library. This study reported empirical findings in peer-reviewed journals from 2015 to 2021. The Scopus database was used to search for relevant studies. The keywords and word combinations are “social presence” and “social presence and e-resources”. Only studies published in the English language were reviewed. A two-technique literature search strategy [65] was used in this study to retrieve social presence elements for the university`s library interface. The Technique I used to identify unique elements of social presence from literature review and web content analysis in twelve universities’ library interfaces (see Table II). Website content analysis [66] was used to investigate elements that offer many unique features to support the implementation of social presence through the e-resources website interface.

TABLE II. NAME OF UNIVERSITY LIBRARY

No	Country	Library Name
1	Australia	Queensland University Library
2	England	Lancaster University Library
3	England	Lincoln University Library
4	Germany	Gothe Universitat Library
5	Japan	Hokkaido University Library
6	Japan	Osaka University Library
7	Malaysia	Tun Seri Lanang Library
8	Netherland	Radboud University Library
9	Netherland	University of Groningen Library
10	Taiwan	National Chang Kung University Library
11	USA	Arizona State University Library
12	USA	Louisiana State University Library

Furthermore, Technique II is used to communicate with five library staff to validate the implementation of social presence elements in the library interface. In this activity, the online interview method was conducted with the library staff hosting a live chat on the library's web. Before that, the researchers introduced themselves and tried to inquire about the e-resources services at the library. Researchers tried to use elements provided in the live chat interface, such as emoticons, attachments, and pictures. Throughout the interview session, the researchers recorded chat recordings through this text. At the end of the interview session, the researcher also had a brief discussion with the staff to improve the information obtained during the interview session. Data obtained from interviews and chat recordings will be analyzed by coordinated based on web content analysis. In this technique, the researchers met online with five university library staff, namely Lancaster library staff, Lincoln Library staff, Radboud Library staff, Tun Seri Lanang Library staff, and Queensland Library staff.

Technique III assesses the need for social presence elements in library users. This activity aimed to validate the social presence element on the e-resources interface in the university library. This test was used to assess social presence elements by comparing elements found from phase I. Then, unique elements identified in the literature review phase were cross-checked with observational results from web content analysis. It was found that the elements searched with library reviews were similar to the search results of social presence elements in web content analysis. There were some additional new elements found after web content analysis was conducted. The findings of social presence elements for e-resources in the university library interface (see Table IV) were then tested with a questionnaire (see Table V). The survey method was performed with 44 users of the e-resources interface in the university library [67] via a google form. The output from this analysis phase is an element of social presence in the design of e-resources interfaces and will be used to design prototypes in the next phase.

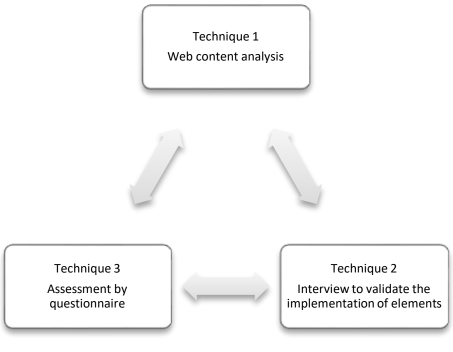

Technique 1
Web content analysis
Technique 3
Assessment by questionnaire
Technique 2
Interview to validate the implementation of elements

Fig. 1. Research methodology.

The three phases used in this study are known as triangulations. This study does not rely on a single source and instead uses a systematic search approach, web content analysis by looking at implemented social presence elements, and interviews with library staff to validate the function of social presence elements in university libraries. Finally, a questionnaire was used to collect data on user requirements.

IV. Results

A. What Elements are Needed for the University Library's e-resources Interface Design?

To answer this question, this study conducted a literature review and web content analysis to answer this question.

1) The need for social presence elements on the interface of e-resources through a literature review: Analysis of the need for elements of social presence on the interface of e-resources is obtained through two stages: literature review and cross-checking through web content analysis of university libraries. As a result of the literature review, the list of elements of social presence was found as follows (see Table III: Table of elements of social presence).

TABLE III. Social Presence Elements

No	Elements
1	Human picture [69]
2	Text [69]
3	Personalized greetings [65]
4	Frequently Asked Questions[65]
5	Instant message [65]
6	Chat [69]
7	Social Networks [65]
8	Product Description [65]
9	Emoticon [69]
10	Colors (Lee et al. 2010)
11	Interaction functions [69]
12	Languages [65]
13	Videos [65]
14	Tag [68]
15	Website data [65]
16	Animation [65]
17	Virtual Agents [65]
18	Number of entries [65]
19	Discussion Forums [65]
20	Graphics [69]
21	Comments [65]
22	Advertisements [65]
23	Ranking [68]
24	User reviews [68]
25	Voice chat [60]
26	Share [28]
27	Avatar [28]
28	Ratings [68]
29	Recommendation system [65]

2) Web content analysis: The stage used in this method is, firstly: surveying whether the design of the university library interface possessed the characteristics or elements of social presence as mentioned in the literature review. Further observations were made to observe the social presence element in the design of the e-resources interface of university libraries from several countries. Then, a web content analysis method [68] was performed to obtain more detailed data related to the need for social presence elements in the design of e-resources interfaces in university libraries.

Based on the study, it is known that out of the twelve university library interface designs that have been observed, only seven libraries have social presence element features namely Lancaster University Library, Lincoln University Library, NCKU library, Radboud University Library, the Tun Seri Lanang Library, the University of Queensland Library, and the Arizona State University Library. Based on web content analysis observation, some elements are found in this method (see Table IV).

TABLE IV. SOCIAL PRESENCE ELEMENTS BASED ON WEB CONTENT ANALYSIS

No	Elements
1	Notification
2	Library account
3	Application/Request
4	Link on reference management
5	Orchid account link
6	Submit an article
7	Publisher chat
8	Pay
9	Forum Group
10	Disabled
11	One search
12	Download history
13	Altmetrics
14	Language
15	Tools
16	Order
17	Borrow
18	Delivery
19	Drive-thru
20	Library Guide
21	Gift
22	Promotion
23	Events
24	Digital library card
25	Attachment
26	Reading list

Then, the following sections will describe the role of each group of social presence elements and their constituent sub-elements in the e-resources interface design in the university library. Table V describe the four main social presence elements suggested to be critical for e-resources in the university’s library interface design are live chat by adding video to create more of a presence and impression, and Library account by adding elements: like, share, rating, and pay. Links to reference manager links to discussion forums to share experiences, links to social media, and links for Accessibility. Table V presents a categorization and description of the identified groups of social presence elements and their constituting sub-elements.

TABLE V. SOCIAL PRESENCE ELEMENTS SUGGESTED FOR E-RESOURCES IN THE UNIVERSITY`S LIBRARY INTERFACE DESIGN

Main element	Sub-element
Live Chat	Human Picture, Text, Chat, Emoticon, Interaction Function, Video, Voice, Virtual Agent, Comment, User Review, Share, Rating, Like, Attachment.
Library Account	Personalized Wellcome, Notification, FAQ, Order Now, User Request, Language Choice, Recommender System, Charges, Read/ Download History, Tool, Booking Collection, Delivery Service, Drive Thru, Library Guide, Gift, Event, Digital Library Card.
Links	Social Network, Discussion Forum, Reference Manager, Orcid.
Accessibility	Font Size, Keyboard Navigation, Readable Font, Color

Furthermore, based on the results of observation and analysis of web content, the role of each group of social presence elements and its constituent sub-elements in the design of e-resources interfaces in university libraries are:

a) Live chat, a finding in the literature review where this critical element is one of the indicators of the existence of social activities that indicate a social presence in a system. Libraries initially used this live chat to make it easier for users to obtain referral services. Live chat services are unique to human nature [69] and have positive word-of-mouth effects [70] and things that support live communication to get help from library staff. So, the elements for live chat are chat, staff name, staff photo, voice, text, emoticons, help, and video, adding video to create more impressions of presence. This element is used to meet and communicate with staff or librarians to get services directly. That way, users expect face-to-face service even if the service is online.

b) A library account that is trustworthy and easy to manage includes various personal information, activities, and daily tasks of library users like user accounts, personalized responses, digital library cards, orders, charges, delivery services, driveways, and download history. This element is used to manage users' personal needs in e-resources services in university libraries by adding user accounts, personalized responses, digital library cards, collection ordering services, late billing, charges, drive-through services, delivery services, download history, recommender system [71], QR, like, share, rating, pay and cart. Users can share videos of new findings from research, experimental videos, videos of researcher observations, and photos of research results. This element of sharing will encourage users to engage [72] in e-resources services at the library.

c) Links are used to improve system performance and make it easier for library users to connect to various networks that support the work of library users. Those elements are links to orcids, referral managers, forum discussions to share experiences, and links to social media.

d) Accessibility is things that make it easier for people with disabilities to access e-resources services. Interface design [73] is essential to consider accessing information created by social and environmental contexts.

With live chat elements, library accounts, links, and accessibility, the e-resources interface in the university library is expected to be fun and convenient for users. Fun and convenience will increase the use of e-resources, and as a result, university performance will increase. Users with disabilities will also find it easier to use e-resources in the library. Users will feel comfortable, happy, and engaged with this service so that the cognitive and affective issues of users when using the interface in the library will be resolved.

3) Interviews: This method is used to ensure that the social presence elements found in the design of the university library interface work well. For that, the researchers try to use these elements to validate the implementation of social presence elements in the library interface.

Based on interviews conducted at the Lancaster library with staff on duty, it was found that the chat element in the Lancaster library had elements of staff names, text, and voice but still needed elements of library staff photos, emoticons, and videos.

In the Lincoln library, there are elements of staff names and text. In contrast, in the Radbud library, there are elements of librarian pictures, voices, attachments, and emoticons on this chat element. Besides that, in Queensland libraries, there are elements of staff names and text.

Based on these interviews, a social presence element in the design of e-resources interfaces in university libraries still needs to be improved. Therefore, in this study, we will further design e-resources interfaces according to data acquisition from library reviews and web content analysis.

B. *Assessment of the Elements of Social Presence based on user Requirement*

The assessment aims to validate the social presence element in the design of the e-resources interface in the university library. This assessment uses a questionnaire with a seven Likert scale to getting more detailed answers from respondents.

Based on the assessment results with the questionnaire, it is known that respondents of this assessment are 43% male and 57% female. Based on their occupations, it is known that 11% are lecturers, 86% are students, and 2% are research assistants. Meanwhile, their level of studies is 45% bachelor, 16% master and 39% Ph.D (see Table VI).

Moreover, based on the respondent's country (see Fig. 2), it is seen that 50% are from Iraq, 14% from Malaysia and Indonesia, 5% from Taiwan, the UK, Libya, and 2% from Thailand, Palestine, Singapore, and Yemen.

Based on the questionnaire results about live chat elements (see Fig. 3), it is known that users need comment element as much as 28.75%, rating element as much as 27.5%, chat element and text element as much as 23.75% and voice element as much as 22.5%.

Based on the study about library account elements (see Fig. 4), it is known that users need a library account element as much as 32.5% for a language choice element, 27.5% for personalized welcome element and tool element.

Based on this study about link elements (see Fig. 5), it is known that users need link elements as much as 27.5% for discussion forum elements, 26.25% for reference manager and orcid link elements. In comparison, users also need social network elements as much as 21.25%.

Based on the study about accessibility (see Fig. 6) , it is known that users need accessibility elements as much as 31.25% for readable font elements, 27.5% for color element, 26.25% for font size elements, and 25% for keyboard navigation.

TABLE VI. RESPONDENT`S DATA

	Category	**Frequency (n=44)**	**Percent (%)**
Gender	Male	19	43
	Female	25	57
Occupation	Lecturer	5	11
	Student	38	86
	Research Assistant	1	2
Level of study	Bachelor	20	45
	Master	7	16
	Ph.D	17	39

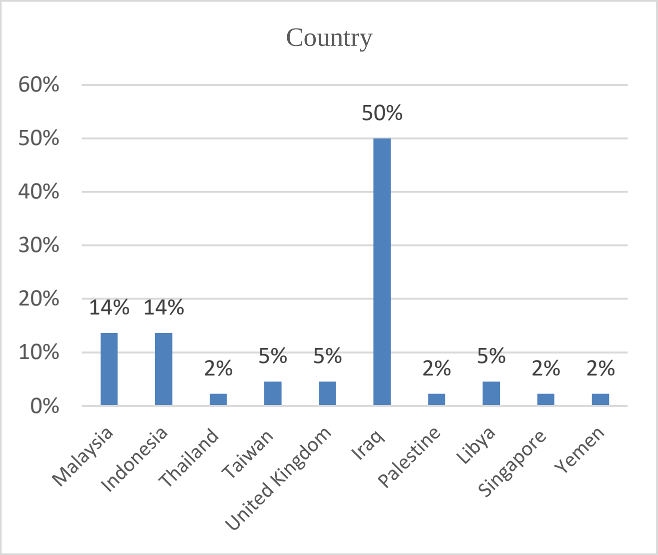


Fig. 2. Respondent's country.

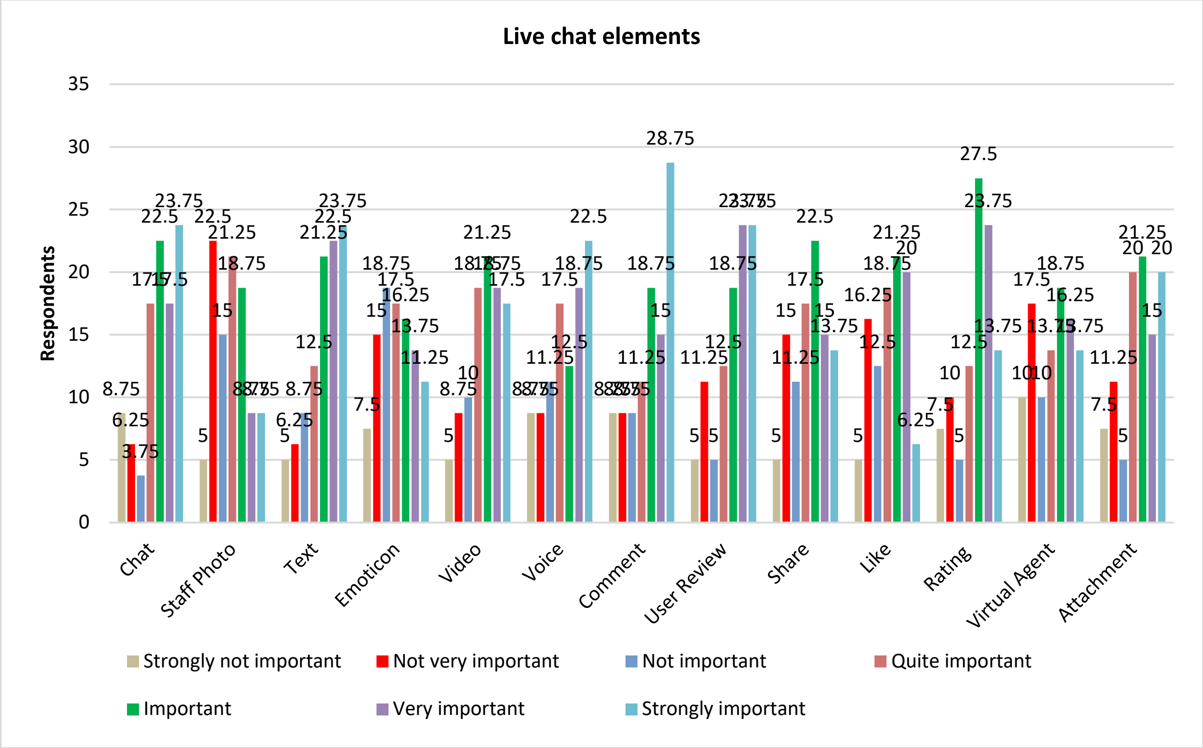


Fig. 3. Live Chat elements.

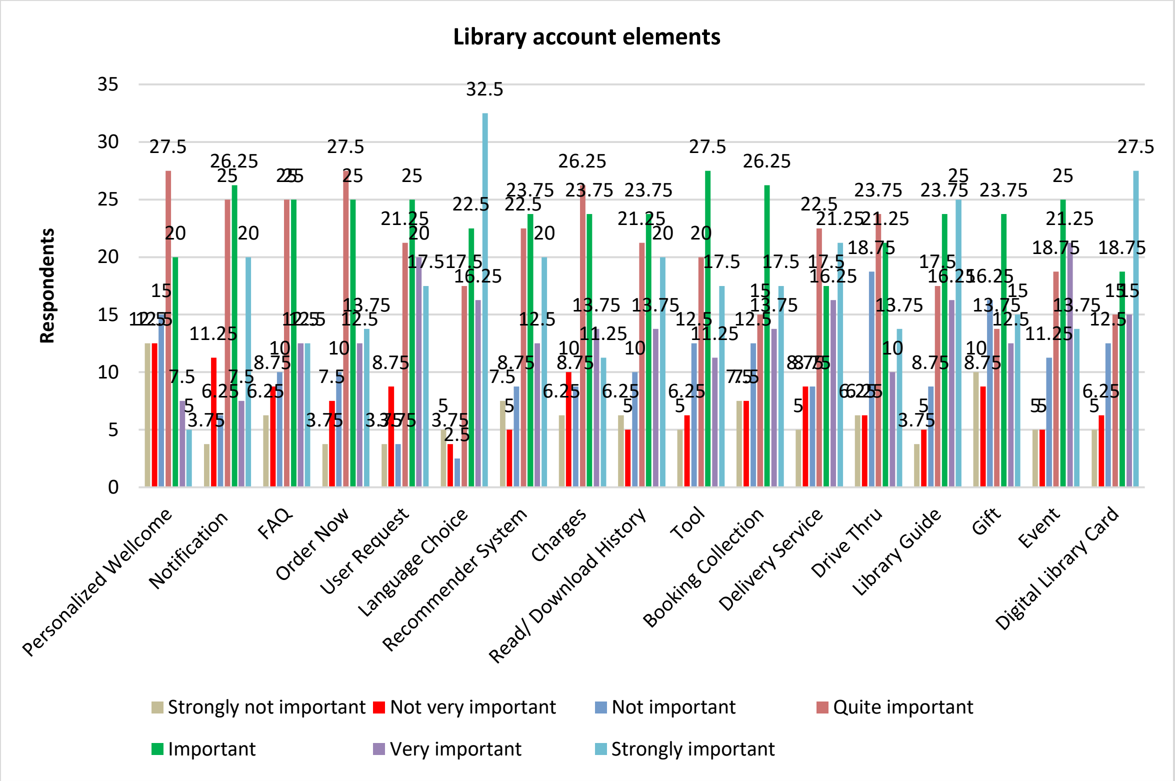


Fig. 4. Library account elements.

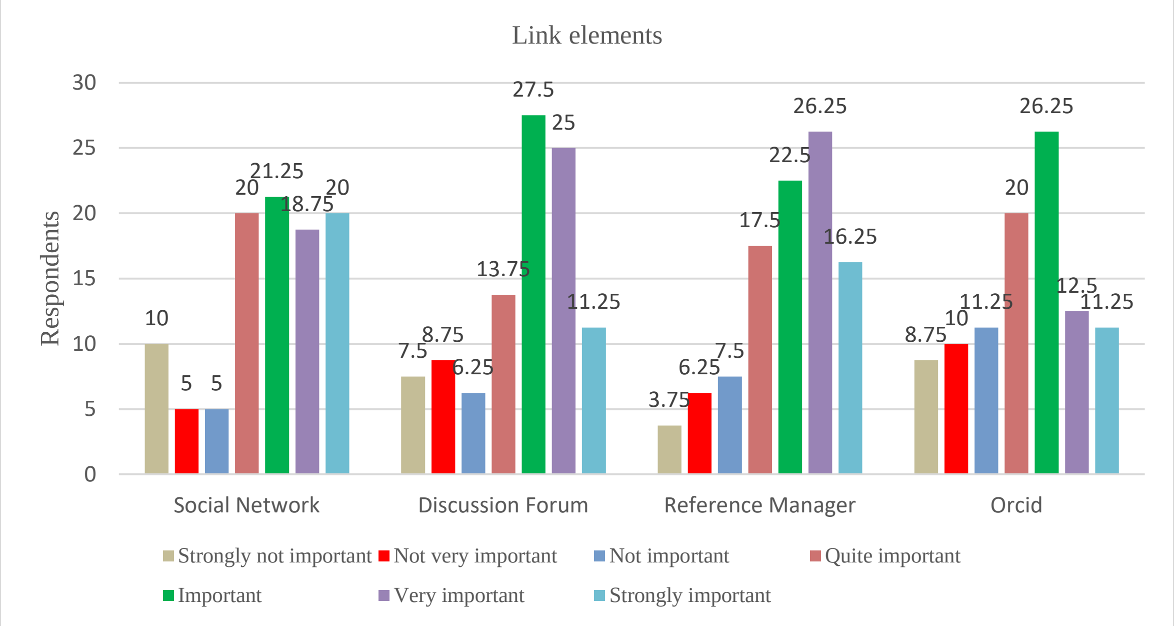


Fig. 5. Link elements.

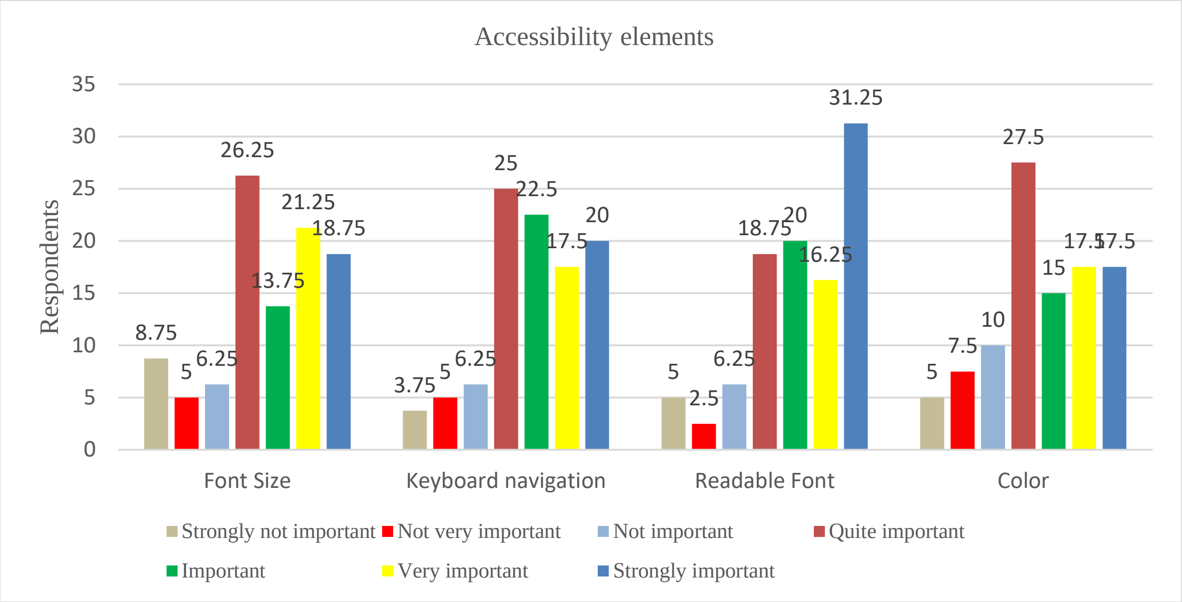


Fig. 6. Accessibility elements.

V. Discussion

This study has looked into twelve university library interfaces in several countries, and the result is that seven of these libraries have an element of social presence. Social presence elements were obtained from a literature review, library interface observations, and observations on electronic commerce. This element in electronic commerce was taken with the consideration that it could be a new element in the design of the e-resources interface in the library.

Researchers in various fields, such as e-learning, e-commerce, and e-services, mainly study social presence. Many studies on social presence show that social presence has a significant effect on interaction, increased learning, and motivation. The social presence model has three categories: emotional expression, open communication, and group cohesion [63]. Emotional expression is shown with humor and self-disclosure. Elements that can represent this category, such as emoticons, videos, or images and open communication, a reciprocal exchange and mutual respect in an interaction. An example of this category is the chat element. Based on a questionnaire data, users need chat elements. This data further supports the problem of difficulties for e-resource users in university libraries. Users need this element to make it easier for them to get help from librarians.

Indeed, social presence is defined as a sense of belonging and acceptance within a group and creating camaraderie among those groups. Social presence can increase education [74], participation, interaction, activity, motivation, learning, sense of presence, and the effectiveness of critical thinking [73]. Then social presence can also increase the use of e-resources in university libraries. Social presence is an important part of communication [75] [76], whether within the individual, in the community, or even in the library. Librarians demonstrate their presence in several ways: in person, over the phone, or on the internet.

Although there are limitations and obstacles in projecting social presence through technology, it can be done well if the element of social presence is raised according to the user requirement of the system. Social presence, which can encourage learning interactions in online learning environments, is considered essential for social learning. When a person perceives high social cues from others, they will gain a better perception of social presence to facilitate learning interactions in online libraries. e-Resources and distance education authorities can use social attendance theory [62] in course planning, applying the principles and creating a friendly learning environment that can increase attendance.

Some considerations obtained from users for the live chat element to improve communications with librarians are (1) proactive and interactive services. Proactive and interactive means users of e-resources that need an immediate response from the librarian to be able to help them or enable live chatting that provides the staff's name, add a selection of common questions most people ask in live chat, and can add some new ideas. This first suggestion directs the conversation and anthropomorphic (need to attribute human characteristics to a particular object). Users need a new chat. They also need Live Agent and mobile live (2) Date and time, which means that users need information about the date and time when using the collection when borrowing facilities from the library. (3) Rating or the recorded number to follow up on some issue. (4) Social Media like Telegram, Twitter, and Google Meet. (5) A unique mark appears if the message has been read (6) login notifications (7) support (8) help (9) updates according to user feedback and [77] QR code.

Another consideration about library accounts, users need (1) date due or expired dates/ renewal notification, (2) rent/ prices, (3) Search (4) Payment through a unique program. In the library or payment upon receipt of request, (5) things matter, (6) Book rental, (7) Acquisition Section, (8) Registration Section (9) Cataloging and Classification, (10) Book storage place, (11) Digital book (12) and library community.

For link elements, users need (1) a professional blog, (2) Most Read, (3) Social communication, (4) A hyperlink points to a whole document or a specific element within a document, (5) HTML with hyperlinks, (6) archive links. Several specialized bodies produce specific software to design unique systems for automatic programming in the field of education so that it is easier for school and college teachers to prepare and present different lessons to their students. These lessons include training, introducing new material, conducting a test, or simulating a specific reality or other existing activities in the classroom. This facilitates and helps the user and spreads to all students through the university library.

For Accessibility, users may add some suggestions. They are (1) sound navigation, (2) Voice over text/ Voice reading, (3) Virtual voice assistance as guidance for e-resources/ Voice search, (4) Fonts type, (5) Color background writing and paper frames as well as numbering, (6) Service Request, (7) assign a supervisor, (8) Navigation Move, (9) A specialized librarian to register users of services available to people with disabilities, (10) sick cart. According to the World Health Organization (WHO), disability has three dimensions: Impairment in a person's body structure or function or mental functioning; examples of impairments include loss of a limb, loss of vision, or memory loss. Thirdly is the limitation of activity [78], such as difficulty seeing, hearing, walking, or problem-solving.

According to social presence theory, information sharing will only be limited if the quality of communication is high. Social presence theory states that creating a cognitive, social presence is relatively challenging because it requires a longer time and more frequent and intensive social interactions. For example, students like to go to the library to study together but also need social activities such as hanging out or shopping. In addition, an effective social presence can only partially be formed, as some students hide from online lessons by not turning on, facing their cameras, or even being present in live sessions. The research in [49] states that some students prefer to watch only recorded lessons.

Social presence is defined as the experience of meeting other people in a medium and interacting with others. Social presence plays a vital role in an online service environment, where there is no face-to-face interaction between the user and the librarian. Previous studies have used social presence theory to explain the social perspective of social commerce. Social presence describes the ability of communication media to convey social signals, such as socially rich messages, virtual agents [51] [79], human-like interfaces, 3D displays [80] [81], and telepresence [48]. Suppose the theory of social presence is applied in the library system. In that case, it will be able to engage library users with the library system or e-resources.

Social presence through interaction between users in e-resources refers to the ability of websites to convey a sense of sociability and human warmth [82] [83]. Web pages containing rich social messages can express a sense of personal presence. The availability of reviews and recommendation features on e-resources pages also increases presence through user interaction.

In an online environment, interacting with photos or videos can create a sense of warmth and human friendliness. Visual presence, such as photos or videos, is essential in attendance through interaction between users and librarians because these elements can describe real people in an online environment. Visual design factors such as images, colors, and graphics can significantly influence people's trust in the online environment. Kracher et al. (2003) also stated that photos could build trust between users and librarians. These visual elements [80], such as images and colors, are essential to the social presence.

VI. Conclusions

Users need help accessing the e-resources system interface in university libraries. These difficulties trigger users' negative emotions and result in the low usage of e-resources in university libraries which causes a decline in university performance. This study contributes to the point of view of increasing the use of e-resources in university libraries by finding elements that can increase positive emotions in users so that users enjoy using e-resources in libraries which will improve the quality of students and lecturers as well as university education and performance. The new elements for the e-resources interface are live chat with voice element, comment, rating and tool. Library account with personalized welcome, preference language, links for reference managers, and social media, and accessibility such as readable help font, color, and font size.

A. Limitation

The limitation of this study is that convenience sampling is used in this study's data collection to facilitate research. More comprehensive data collection techniques can be used in future studies to get more detailed data.

B. Future Work

For future research, the research method needs to be strengthened by interviews with the users to get their views. Secondly, user experience studies will be carried out using the library's e-resources system. Third, user modelling will be carried out to obtain data on an actual user's needs.

References

[1] C. Ganoe, J. M. Carroll, and H. Jiang, "Four requirements for digital case study libraries," Educ. Inf. Technol., vol. 15, pages2, 2010.

[2] B. K. Anhwere and A.-A. Paulina, "Accessibility and postgraduate students use of electronic resources in university of Cape Coast," Res. J. Libr. Inf. Sci., vol. 2, no. 1, pp. 9–14, 2018, [Online]. Available: https://www.sryahwapublications.com/research-journal-of-library-and-information-science/pdf/v2-i1/2.pdf.

[3] P. Handayani et al., "The Evaluate off Usability Web Design Based on the User Experience," J. Phys. Conf. Ser., vol. 1779, no. 1, p. 012012, 2021, doi: 10.1088/1742-6596/1779/1/012012.

[4] M. Wójcik, "How to design innovative information services at the library?," Libr. Hi Tech, vol. 37, no. 2, pp. 138–154, 2019, doi: 10.1108/LHT-07-2018-0094.

[5] N. F. Taharim, N. K. Zainal, and W. X. Lim, "An affective design guideline to optimize higher institution websites," Adv. Intell. Syst. Comput., vol. 739, pp. 771–780, 2018, doi: 10.1007/978-981-10-8612-0_80.

[6] Y. H. Chen and I. Chengalur-Smith, "Factors influencing students' use of a library Web portal: Applying course-integrated information literacy instruction as an intervention," Internet High. Educ., vol. 26, pp. 42–55, 2015, doi: 10.1016/j.iheduc.2015.04.005.

[7] A. Fry and L. Rich, "Usability Testing for e-Resource Discovery: How Students Find and Choose e-Resources Using Library Web Sites," J. Acad. Librariansh., vol. 37, no. 5, pp. 386–401, 2011, doi: 10.1016/j.acalib.2011.06.003.

[8] A. Kundu, "Usage of E-Resources among Law Students in NUJS Library," vol. 10, no. 1, p. 6, 2021.

[9] H. de Ribaupierre and G. Falquet, "Extracting discourse elements and annotating scientific documents using the SciAnnotDoc model: a use case in gender documents," Int. J. Digit. Libr., vol. 19, no. 2–3, pp. 271–286, 2018, doi: 10.1007/s00799-017-0227-5.

[10] Á. Tejeda-Lorente, C. Porcel, E. Peis, R. Sanz, and E. Herrera-Viedma, "A quality based recommender system to disseminate information in a university digital library," Inf. Sci. (Ny)., vol. 261, pp. 52–69, 2014, doi: 10.1016/j.ins.2013.10.036.

[11] P. Wijetunge, "Usage of electronic resources by librarians of sri lankan universities," Ann. Libr. Inf. Stud., vol. 64, no. 1, pp. 21–27, 2017.

[12] N. K. Soni, S. Rani, A. Kumar, and J. Shrivastava, "Evaluation of usage of e-resources and inmas library services through user's perspective: An analytical study," DESIDOC J. Libr. Inf. Technol., vol. 40, no. 4, pp. 238–246, 2020, doi: 10.14429/djlit.40.4.16047.

[13] T. Sritharan, "Evaluation of Usage and User Satisfaction on Electronic Information Resources and Services: A Study at Postgraduate," J. Univ. Libr. Assoc. Sri Lanka, vol. 21, no. 2, pp. 73–88, 2018.

[14] N. Nordin and F. Hassan, "Student Perception on the use of Tablet Computer in Academic Library," Asia-Pacific J. Inf. Technol. Multimed., vol. 07, no. 01, pp. 45–56, 2018, doi: 10.17576/apjitm-2018-0701-04.

[15] R. N. Bellary and S. Surve, "E-Resources are boon for the teaching and research work of an academic institute: A survey on usage and awareness of e-resources by the NMIMS (Deemed University) engineering faculties, Mumbai," Libr. Philos. Pract., vol. 2019, 2019.

[16] A. S. Katabalwa, "Use of electronic journal resources by postgraduate students at the University of Dar es Salaam," Libr. Rev., vol. 65, no. 6–7, pp. 445–460, 2016, doi: 10.1108/LR-11-2015-0108.

[17] A. Tella, F. Orim, D. M. Ibrahim, and S. A. Memudu, "The use of electronic resources by academic staff at The University of Ilorin, Nigeria," Educ. Inf. Technol. Vol., vol. 23, pp. pages9–27, 2018.

[18] Y. Li and C. Liu, "Information Resource, Interface, and Tasks as User Interaction Components for Digital Library Evaluation," Inf. Process. Manag., vol. 56, no. 3, pp. 704–720, 2019, doi: 10.1016/j.ipm.2018.10.012.

[19] C. Chukwueke, "Availability of e-Resources and Accessibility of e-Services in Academic and Special Libraries in Abia State, Nigeria," J. Libr. Inf. Science Technol., no. May, 2017, [Online]. Available: http://www.iaeme.com/issue.asp?JType=JLIST&VType3&IType=1.

[20] E. Lwoga and F. Sukums, "Health sciences faculty usage behaviour of electronic resources and their information literacy practices," Glob. Knowledge, Mem. Commun., vol. 67, no. 1–2, pp. 2–18, 2018, doi: 10.1108/GKMC-06-2017-0054.

[21] K. A. Eiriemiokhale, "Frequency of Use and Awareness of Electronic Databases By University Lecturers in South-West, Nigeria," Libr. Philos. Pract., vol. 2020, pp. 1–23, 2020.

[22] R. J. Garg, V. Kumar, and Vandana, "Factors affecting usage of e-resources: scale development and validation," Aslib J. Inf. Manag., vol. 69, no. 1, pp. 64–75, 2017, doi: 10.1108/AJIM-07-2016-0104.

[23] B. Massis, "The user experience (UX) in libraries," Inf. Learn. Sci., vol. 119, no. 3–4, pp. 241–244, 2018, doi: 10.1108/ILS-12-2017-0132.

[24] X. Wang, J. Li, M. Yang, Y. Chen, and X. Xu, "An empirical study on the factors influencing mobile library usage in IoT era," Libr. Hi Tech, vol. 36, no. 4, pp. 605–621, 2018, doi: 10.1108/LHT-01-2018-0008.

[25] D. P. Srirahayu, "User Analysis of Library Usage to Fulfill Information Needs," Khizanah al-Hikmah J. Ilmu Perpustakaan, Informasi, dan Kearsipan, vol. 7, no. 2, p. 115, 2019, doi: 10.24252/kah.v7i2a2.

[26] N. B. P. Edem, "Availability and Utilization of Electronic Resources by Postgraduate Students in a Nigerian University Library : A Case Study of University of Calabar , Nigeria," vol. 6, no. 2, pp. 60–69, 2016.

[27] K. Hill, "Usability beyond the Home Page: Bringing Usability into the Technical Services Workflow," Ser. Libr., vol. 78, no. 1–4, pp. 173–180, 2020, doi: 10.1080/0361526X.2020.1702857.

[28] Y. J. Kang and W. J. Lee, "Effects of sense of control and social presence on customer experience and e-service quality," Inf. Dev., vol. 34, no. 3, pp. 242–260, 2018, doi: 10.1177/0266666916686820.

[29] T. S. M. Tengku Wook et al., "User Experience Evaluation Towards Interface Design of Digital Footprint Awareness Application," Asia-Pacific J. Inf. Technol. Multimed., vol. 09, no. 01, pp. 17–27, 2020, doi: 10.17576/apjitm-2020-0901-02.

[30] J. T. Bickle, M. Hirudayaraj, and A. Doyle, "Social Presence Theory: Relevance for HRD/VHRD Research and Practice," Adv. Dev. Hum. Resour., vol. 21, no. 3, pp. 383–399, 2019, doi: 10.1177/1523422319851477.

[31] K. Blessinger and D. Comeaux, “User experience with a new public interface for an integrated library system,” Inf. Technol. Libr., vol. 39, no. 1, pp. 1–18, 2020, doi: 10.6017/ITAL.V39I1.11607.

[32] L. Sejane, “Access to and use of electronic information resources in the academic libraries of the lesotho library consortium,” 2017.

[33] S. Thanuskodi and A. Ashok Kumar, “Usage of electronic resources among ophthalmologists,” Libr. Philos. Pract., vol. 2017, no. 1, 2017.

[34] M. Mani, A. Thirumagal, B. Vijayalakshmi, and E. Priyadharshini, “Usage of E-Resources among the students of South Tamil Nadu with the Special Reference of Manonmaniam Sundaranar University, Tirunelveli - A study,” Libr. Philos. Pract., vol. 2019, 2019.

[35] M. Hussin, M. S. Said, N. Mohd Norowi, N. A. Husin, and M. R. Mustaffa, “Authentic Assessment for Affective Domain Through Student Participant in Community Services,” Asia-Pacific J. Inf. Technol. Multimed., vol. 10, no. 01, pp. 52–62, 2021, doi: 10.17576/apjitm-2021-1001-05.

[36] J. Kim, K. Merrill, and H. Yang, “Why we make the choices we do: Social TV viewing experiences and the mediating role of social presence,” Telemat. Informatics, vol. 45, no. August, p. 101281, 2019, doi: 10.1016/j.tele.2019.101281.

[37] D. Narciso, M. Bessa, M. Melo, A. Coelho, J. Vasconcelos-Raposo, and M. Čertický, “Immersive 360∘ video user experience: impact of different variables in the sense of presence and cybersickness,” Univers. Access Inf. Soc., vol. 18, pp. pages77–87, 2019.

[38] S. Schneider, M. Beege, S. Nebel, L. Schnaubert, and G. D. Rey, The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE). Educational Psychology Review, 2021.

[39] Y. M. Aldheleai, Z. Tasir, W. M. Al-Rahmi, M. A. Al-Sharafi, and A. Mydin, “Modeling of students online social presence on social networking sites with academic performance,” Int. J. Emerg. Technol. Learn., vol. 15, no. 12, pp. 56–71, 2020, doi: 10.3991/ijet.v15i12.12599.

[40] G. L. Mallmann and A. C. G. Maçada, “Shadow IT and CompuTer-medIaTed CollaboraTIon: developIng a Framework baSed on SoCIal preSenCe Theory,” Rev. Adm. UFSM, St. maria, vol. 12, no. 4, pp. 821–839, 2019, doi: DOI: 10.5902/19834659 23853.

[41] J. A. Alzahrani, “Use and Impact of Electronic Resources At King Abdulaziz University , Jeddah , Saudi Arabia,” vol. 9, no. 4, pp. 60–66, 2019.

[42] M. Rafi, Z. JianMing, and K. Ahmad, “Technology integration for students’ information and digital literacy education in academic libraries,” Inf. Discov. Deliv., vol. 47, no. 4, pp. 203–217, 2019, doi: 10.1108/IDD-07-2019-0049.

[43] Y. Rohmiyati, T. S. M. T. Wook, and N. Sahari, “The Usage of Electronic Resources in Libraries,” 2021.

[44] G. Cao, M. Liang, and X. Li, “How to make the library smart? The conceptualization of the smart library,” Electron. Libr., vol. 36, no. 5, pp. 811–825, 2018, doi: 10.1108/EL-11-2017-0248.

[45] H. Mustafa, A. Mohammad, D. Iyad Abu Abu, A. Gheed Mufied, A.-A. Fatima Abdalla, and A. Mouhammd Mahmoud, “Evaluating Usability and Content Accessibility for e-Learning Websites in the Middle East,” Int. J. Technol. Hum. Interact. 16(1) DOI 10.4018/IJTHI.2020010104, 2020.

[46] Y. Udjaja, Sasmoko, Y. Indrianti, O. A. Rashwan, and S. A. Widhoyoko, “Designing Website E-Learning Based on Integration of Technology Enhance Learning and Human Computer Interaction,” 2018 2nd Int. Conf. Informatics Comput. Sci. ICICoS 2018, pp. 71–74, 2018, doi: 10.1109/ICICOS.2018.8621792.

[47] T. W. Liew, S. M. Tan, and H. Ismail, “Exploring the effects of a non-interactive talking avatar on social presence, credibility, trust, and patronage intention in an e-commerce website,” Human-centric Comput. Inf. Sci., vol. 7, no. 1, 2017, doi: 10.1186/s13673-017-0123-4.

[48] R. Algharabat, N. P. Rana, Y. K. Dwivedi, A. A. Alalwan, and Z. Qasem, “The effect of telepresence, social presence and involvement on consumer brand engagement: An empirical study of non-profit organizations,” J. Retail. Consum. Serv., vol. 40, no. July 2017, pp. 139–149, 2018, doi: 10.1016/j.jretconser.2017.09.011.

[49] W. Jing, “Person - to - person interactions in online classroom settings under the impact of COVID - 19 : a social presence theory perspective,” Asia Pacific Educ. Rev., no. 0123456789, 2021, doi: 10.1007/s12564-021-09673-1.

[50] C. S. Oh, J. N. Bailenson, and G. F. Welch, “A systematic review of social presence: Definition, antecedents, and implications,” Front. Robot. AI, vol. 5, no. OCT, pp. 1–35, 2018, doi: 10.3389/frobt.2018.00114.

[51] L. Finley, “Big Picture Presence: Bringing Teaching Presence to the Forefront,” 2017, [Online]. Available: https://scholarspace.jccc.edu/cgi/viewcontent.cgi?article=1228&context=c2c_sidlit%0Ahttp://scholarspace.jccc.edu/cgi/viewcontent.cgi?article=1228&context=c2c_sidlit.

[52] S. Molinillo, R. Aguilar-Illescas, R. Anaya-Sánchez, and M. Vallespín-Arán, “Exploring the impacts of interactions, social presence and emotional engagement on active collaborative learning in a social web-based environment,” Comput. Educ., vol. 123, no. April, pp. 41–52, 2018, doi: 10.1016/j.compedu.2018.04.012.

[53] E. N. Anyaoku and L. O. Akpojotor, “Usability Evaluation of University Library Websites in South-South Nigeria,” Libr. Philos. Pract., vol. 2020, no. January, pp. 1–26, 2020.

[54] N. R. Zulkifli, N. Sahari, N. Azan, M. Zin, and R. A. Majid, “Inclusive Design Requirement in Designing Accessibility for Low Cognitive Users Keperluan Reka Bentuk Inklusif dalam Reka Bentuk Ketercapaian untuk Pengguna Kognitif Rendah,” vol. 12, no. 1, pp. 1–12, 2023.

[55] K. D. Abbas and U. M. Song, “Accessibility and Utilization of Electronic Information Resources for Research Activities in Agricultural Research Institutes in Kaduna State, Nigeria,” Covenant J. Libr. Inf. Sci., vol. 3, no. 1, pp. 1–11, 2020, doi: 10.47231/skjc6572.

[56] G. N. Kamau, K. Jomo, and N. Dorothy, “An Assessment of The Accessibility of Electronic Information Resources by Academic Library Users: A Case of The University of Nairobi,” Moi Univ. Press, 2017.

[57] J. Idiegbeyan-Ose, G. Ifijeh, A. Aregbesola, S. Owolabi, and E. Toluwani, “E-resources vs prints: Usages and preferences by undergraduates in a private university, Nigeria,” DESIDOC J. Libr. Inf. Technol., vol. 39, no. 2, pp. 125–130, 2019, doi: 10.14429/djlit.39.2.13885.

[58] F. Gatwiri Kiambati, “Web Accessibility and Use of Assistive Technology in Accessing E-Resources by Learners with Visual Impairments,” East African J. Inf. Sci., 2018, doi: 10.21428/31568843.

[59] J. Song, H. Moon, and M. Kim, “When do customers engage in brand pages? Effects of social presence,” Int. J. Contemp. Hosp. Manag., vol. 31, no. 9, pp. 3627–3645, 2019, doi: 10.1108/IJCHM-10-2018-0816.

[60] J. Wei, S. Seedorf, P. B. Lowry, C. Thum, and T. Schulze, “How increased social presence through co-browsing influences user engagement in collaborative online shopping,” Electron. Commer. Res. Appl., vol. 24, pp. 84–99, 2017, doi: 10.1016/j.elerap.2017.07.002.

[61] G. D. Voinea et al., “Study of Social Presence While Interacting in Metaverse with an Augmented Avatar during Autonomous Driving,” Appl. Sci., vol. 12, no. 22, 2022, doi: 10.3390/app122211804.

[62] K. Aliabadi and M. Zare, “The social presence theory in distance education; the role of social presence in web-based educational environment,” FMEJ 7;4 mums.ac.ir/j-fmej December 23, 2017, pp. 53–54, 2017.

[63] J. Richardson and P. Lowenthal, “Social presence in online learning: multiple perspectives on practice and research,” Soc. Presence Online Learn. Mult. Perspect. Pract. Res., pp. 86–98, 2017.

[64] J. M. Basch, K. G. Melchers, J. Kegelmann, and L. Lieb, “Smile for the camera! The role of social presence and impression management in perceptions of technology-mediated interviews,” J. Manag. Psychol., vol. 35, no. 4, pp. 285–299, 2020, doi: 10.1108/JMP-09-2018-0398.

[65] J. Chi, W. Pian, and S. Zhang, “Consumer health information needs: A systematic review of instrument development,” Inf. Process. Manag., vol. 57, no. 6, p. 102376, 2020, doi: 10.1016/j.ipm.2020.102376.

[66] A. Rahman and M. Sadik Batcha, “Content analysis of library websites of select colleges of Delhi University: A study,” DESIDOC J. Libr. Inf. Technol., vol. 40, no. 4, pp. 247–252, 2020, doi: 10.14429/djlit.40.4.15454.

[67] Jakob Nielsen, Usability Engineering, 1st ed. California: Morgan Kaufmann Publishers Inc.340 Pine Street, Sixth FloorSan FranciscoCAUnited States, 1994.

[68] N. N. Ab Rahaman, T. S. M. Tengku Wook, and N. Sahari@Ashaari, “The Design of Adaptive Hypermedia Interface for Children’s Digital Library,” Asia-Pacific J. Inf. Technol. Multimed., vol. 02, no. 02, pp. 1–12, 2013, doi: 10.17576/apjitm-2013-0202-01.

[69] G. Mclean, A. Wilson, and V. Pitardi, “How live chat assistants drive travel consumers ’ attitudes , trust and purchase intentions The role of human touch,” pp. 1795–1812, 2020, doi: 10.1108/IJCHM-07-2019-0605.

[70] L. Rajaobelina, I. Brun, N. Kilani, and L. Ricard, “Examining emotions linked to live chat services : The role of e - service quality and impact on word of mouth,” J. Financ. Serv. Mark., no. 0123456789, 2021, doi: 10.1057/s41264-021-00119-8.

[71] G. Kortemeyer and S. Dröschler, “A user-transaction-based recommendation strategy for an educational digital library Gerd Kortemeyer & Stefan Dröschler,” Int. J. Digit. Libr., vol. 22, pp. 22, pages147–157 (2021), 2021.

[72] L. Zhang and E. H. Jung, “How does WeChat’s active engagement with health information contribute to psychological well-being through social capital?,” Univers. Access Inf. Soc., no. 0123456789, 2021, doi: 10.1007/s10209-021-00795-2.

[73] I. Xie, R. Babu, T. H. Lee, M. D. Castillo, S. You, and A. M. Hanlon, “Enhancing usability of digital libraries: Designing help features to support blind and visually impaired users,” Inf. Process. Manag., vol. 57, no. 3, p. 102110, 2020, doi: 10.1016/j.ipm.2019.102110.

[74] N.-S. C. & K. Chun-Wang Wei, “A model for social presence in online classrooms,” Educ. Technol. Res. Dev. Vol. 60, pages529–545(2012), 2012.

[75] X. Chen and H. Wang, “Automated chat transcript analysis using topic modeling for library reference services,” Proc. Assoc. Inf. Sci. Technol., vol. 56, no. 1, pp. 368–371, 2019, doi: 10.1002/pra2.31.

[76] I. C. Hsu and C. C. Chang, “Integrating machine learning and open data into social Chatbot for filtering information rumor,” J. Ambient Intell. Humaniz. Comput., no. 0123456789, 2020, doi: 10.1007/s12652-020-02119-3.

[77] A. Shah and R. Bano, “ Smart library: Need of 21 st Century ,” Libr. Prog., vol. 40, no. 1, p. 1, 2020, doi: 10.5958/2320-317x.2020.00001.x.

[78] E.-Y. Park, “Digital competence and internet use/behavior of persons with disabilities in PC and smart device use,” Univers. Access Inf. Soc., 2020.

[79] D. G. M. Schouten, A. A. Deneka, M. Theune, M. A. Neerincx, and A. H. M. Cremers, “An embodied conversational agent coach to support societal participation learning by low-literate users,” Univers. Access Inf. Soc., no. 1, 2022, doi: 10.1007/s10209-021-00865-5.

[80] C. Jiang, R. Muhammad, and J. Wang, “Journal of Retailing and Consumer Services Investigating the role of social presence dimensions and information support on consumers ’ trust and shopping intentions,” J. Retail. Consum. Serv., vol. 51, no. December 2018, pp. 263–270, 2019, doi: 10.1016/j.jretconser.2019.06.007.

[81] M. Jamil, “Pemanfaatan Teknologi Virtual Reality (VR) di Perpustakaan,” Bul. Perpust. Univ. Islam Indones., vol. 1, no. 1, pp. 99–113, 2018, [Online]. Available: https://journal.uii.ac.id/Buletin-Perpustakaan/article/download/11503/8674.

[82] J. Weidlich and T. J. Bastiaens, “Explaining social presence and the quality of online learning with the SIPS model,” Computers in Human Behavior, vol. 72. pp. 479–487, 2017, doi: 10.1016/j.chb.2017.03.016.

[83] S. S. Engku Alwi and T. S. M. Tengku Wook, “Social presence model for e-commerce,” J. Teknol., vol. 77, no. 1, pp. 71–83, 2015, doi: 10.11113/jt.v77.4147.

Multidimensional Private Information Portrait in Social Network Users

Fangfang Shan[1]*, Mengyi Wang[2], Huifang Sun[3]
School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, Henan, China[1, 2, 3]
Henan Key Laboratory of Cyberspace Situation Awareness, Zhengzhou 450001, Henan, China[1]

***Abstract*—In order to tackle the challenges of users' weak privacy awareness and frequent disclosure of private information in social network, this paper proposes a multidimensional privacy information portrait model of users in Chinese social networks. Because the TF-IDF (Term Frequency-Inverse Document Frequency, TF-IDF) algorithm does not consider the distribution of feature terms among and within classes, uses the TF-IDF algorithm based on the bag-of-words model to calculate the sensitivity of user privacy information. Considering the diversity of user privacy information, this paper proposes the PROLM (Positive reverse order lookaround matching) algorithm, which is combined with the Flashtext+ (improved Flashtext) algorithm and SMA (string matching algorithm, SMA), the PROLM_FlashText+_SMA to extract user personal privacy information and location where the privacy information is located, and return the sensitivity. Using the BERT (Bidirectional Encoder Representation from Transformers, BERT)-Softmax privacy information classification model, the privacy information is classified into high, moderate and mild privacy information, and a multidimensional privacy information portrait of the user is constructed based on the privacy information and sensitivity. The experiments show that the accuracy of PROLM_FlashText+_SMA algorithm for privacy information extraction reaches 93.63%, and the overall F1 index of privacy information classification using the BERT-Softmax model reaches 0.9798 on the test set, better than baseline comparison model, has better privacy information classification effect.**

***Keywords*—Social network; personal privacy information; privacy information portrait; sensitivity; privacy protection; BERT**

I. Introduction

In the current stage, social networking platforms like QQ, WeChat, Weibo, and Facebook have spread at an unprecedented speed, becoming indispensable parts of people's lives. According to the 50th Statistical Report on China's Internet Development, As of June 2022, 38% of internet users said they had encountered any online security problems in the past six months. In addition, the proportion of internet users who experienced personal information disclosure was the highest, at 22.1%. The emergence of social networks has changed the way people communicate and interact with each other [1], but people's use of social networks to share their lives inevitably brings the risk of privacy leakage. A particular piece of data posted by a user on a social networking platform may contain one privacy item of the user, but if many pieces of information contain privacy items of the user, these privacy items may be associated to expose a whole privacy chain of the user. For the published information, users cannot control its dissemination path, and once the information is accessed by unlawful elements, there is a risk of causing economic and property losses, and even threatening personal safety [2]. For example, in August 2022, an individual posted on a hacker forum claiming to auction the Shanghai Health Code database for $4000. The post stated that the database contained personal information of 48.5 million users of the Shanghai Health Code, including ID numbers, names, and phone numbers of individuals who have either resided in or visited Shanghai since the implementation of the health code system. The post also included the release of 47 sets of sample data. Such incidents have raised concerns about personal privacy and have led to widespread interest in the study of privacy protection technologies in social networks.

At present, many scholars both domestically and internationally are conducting research on privacy protection schemes for social networks. AL-Asbahi R [3] introduced the concept of structural anonymity as a means to reduce data anonymization. Lian *et al*. [4] achieved privacy protection for users by calculating sensitive attribute levels and using different anonymization methods for different levels, but this algorithm has high time expenditure compared to other algorithms. To meet the differences in privacy protection needs of different users, Yin X et al. [5] proposed a social network attributes graphs algorithm under personalized differential privacy, which is for the independent attribute information between users. Ning et al. [6] integrated noisy weights into the generated graph to solve the problem of edge weights and frequent structure privacy and realized the privacy protection of graph structure in the process of frequent mining. Huang et al. [7] proposed the PBCN method based on the joint clustering and randomization algorithm to resist node attacks and degree attacks in social networks and lost adjacency information. However, the error generated by the method includes noise error and graph reconstruction error.

However, these privacy protection studies have not yet visualized the personal privacy information that users expose on social networking platforms, making it difficult for users to intuitively understand the potential harm caused by such privacy information. Social networking platforms store a large amount of personal privacy information from users, and due to their massive user base and strong communication interactivity, users' privacy information is more prone to leakage. In order to identify potential privacy leakage risks and

This paper is supported by the National Natural Science Foundation of China (No. 62302540), Open Foundation of Henan Key Laboratory of Cyberspace Situation Awareness (No. HNTS2022020); Natural Science Foundation of Henan Province Youth Science Fund Project (No. 232300420422) and the Natural Science Foundation of Zhongyuan University of Technology (No. K2023QN018).

attack threats, enhance users' privacy awareness and behavior, and strengthen personal privacy protection, reduce the probability and impact of privacy breaches, the author extracts users' exposed privacy information from their historical data posted on social networking platforms and uses it to construct a multidimensional privacy information portrait of users.

The main contributions of this paper are as follows:

- Use the BOW-TF-IDF algorithm to measure the sensitivity of privacy information. The BOW-TF-IDF algorithm can better consider the distribution of privacy information within and between classes, more accurately reflect the differences between different categories, and has certain advantages in measuring privacy information sensitivity.
- In response to the limitation of the FlashText algorithm that can only extract English text, this paper improves it to make it suitable for Chinese scenarios. Propose PROLM_FlashText+_SMA algorithm extracts user privacy information. Experiments have shown that this algorithm has advantages in terms of precision in extracting privacy information.
- Build a privacy information classification model. Using the BERT-Softmax model to classify privacy information. This paper uses one hot to encode text, converting unstructured text sequences into structured feature vectors, and inputting text features into the fully connected layer to calculate the probability of each category label through the Softmax function. Experimental analysis shows that this classification model is significantly superior to traditional classification models.
- This paper proposes the concept of user portrait for privacy information, which determines the risk of user privacy leakage by calculating the average sensitivity of privacy information.

This article is mainly divided into four chapters. In Section I, introduction is mentioned which mainly introduces the background and significance of this study. Section II is related work, which introduces the research of domestic and foreign scholars on user portrait. Section III is algorithms and model, which introduces the algorithm and model of this article. Section IV is experimental results and analysis and Section V concludes the paper.

II. Related Work

At present, research on user portrait can be broadly classified into three main categories: user behavior-based user portrait model, interest-based user portrait model, and topic-based user portrait model.

Alan Cooper was the first to suggest the idea of a user portrait. A user portrait is a designated user model based on some actual data points (such as social traits and consumption variables) [8].

1) User behavior-based user portrait model: Li et al. [9] developed a model that utilizes big data mining and analysis technology to construct a student portrait based on extensive data from campuses. By extracting the characteristics and attributes from a vast amount of behavioral data, they are able to create a comprehensive portrait of a student. Minghui You et al. [10] proposed a behavior-aware user profiling technique that utilizes data mining of user attributes to construct an initial user portrait by identifying user behavior patterns through perception. Zhang et al. [11] introduced an enhanced fuzzy MLKNN multi-label learning algorithm based on MLKNN, aiming to address the challenges of subjective augmentation caused by credit data discretization and the absence of multi-dimensional *credit user portrait in current credit data research.*

2) Interest-based user portrait model: Shufang Wu et al. [12] put forward an interest transfer-based user portrait building approach, which serves as a remedy for the inadequacy and inaccuracy of current microblog user portrait creation methods in capturing user characteristics. Ding Z et al. [13] proposed the LDA-RCC model to analyze the interests of forum users and create user portrait.

3) Topic-based user portrait model: Jianyun Wu et al. [14] proposed by analyzing the crawled videos, users, and their viewing data through text mining, this paper build a single user portrait, cluster the users, and extract themes through K-Means and LDA models to explore the characteristics of group users. Deng et al. [15] proposed a user portrait fraud warning scheme based on publicly available data on Weibo. They conducted preliminary screening and cleaning based on the keyword "being scammed" to obtain effective fraudulent user identity documents. Through feature engineering techniques such as avatar recognition, artificial intelligence (AI) sentiment analysis, data filtering, and fan blogger type analysis, these images and texts were abstracted into user preferences and personality characteristics, and multi-dimensional information was integrated to construct user portrait.

With the increasing awareness of privacy among individuals, personal privacy issues are increasingly receiving attention. To address these issues, this paper applies user profiling technology to social networks and proposes the concept of user privacy information profiling for social networks. By extracting five dimensions of privacy information from the historical information shared by users on social networks, a user portrait of privacy information is constructed to visually display potential privacy threats and risks to users, and to intuitively understand potential privacy issues and leakage risks.

To benefit the readers with a quick reference, the major notations of this paper are listed in Table I.

TABLE I. MAJOR NOTATIONS

Notation	Description
TF	Term Frequency
IDF	Inverse document frequency
TF-IDF	sensitivity w_i
w_i	Sensitivity of privacy information
Current (P_i)	Location of the node where the current keyword is located
next(*Current*,P_i^j)	Location of the node where the next keyword is located
F	Transfer Functions
Query(Q)	The relationship between the current word to be queried and other words
Key(K)	Vector used to calculate attention weights
Value(V)	Weighted results
W^Q,W^K,W^V	Weight matrix
d_k	Dimension of input information
W^O	Mapping Vector for Multi head
Concat(·)	Concatenation function
$h_\theta(x^i)$	Softmax regression model discriminant function
J_θ	The Cost Function of Softmax Regression Model
x^i	Input samples

III. ALGORITHMS AND MODEL

The multidimensional privacy information portrait of social network users is mainly divided into three modules: the privacy information sensitivity calculation model, the social network user privacy information extraction model, and the BERT-based privacy information classification model. The overall structure of the Chinese social network user multidimensional privacy portrait model is shown in Fig. 1.

The user portrait of social network privacy information depicts the personal privacy data exposed by users while using social networks from various perspectives and dimensions. Firstly, the word bag model is utilized to extract features related to privacy information. It involves transforming the sparse word frequency matrix regarding privacy information into a word bag vector. Subsequently, the TF-IDF algorithm is employed to determine the sensitivity of the privacy data. Next, the PROLM_FlashText+_SMA algorithm is utilized to extract the user's privacy information and its corresponding sensitivity. The average sensitivity of the user's privacy data is calculated to classify the risk of privacy leakage into high, medium, and low levels. Lastly, the BERT Softmax model is employed to categorize privacy information as high privacy, moderate privacy, or mild privacy, enabling the construction of a user portrait for social network privacy information.

A. *A Model for Calculating the Sensitivity of Privacy Information of Social Network Users*

This paper employs the TF-IDF algorithm, utilizing the word bag model, to assess the sensitivity of private information. The concept of TF-IDF was introduced by Karen Spärck Jones [16] and entails a fusion of term frequency (TF) and inverse document frequency (IDF).

Bag-of-words model, is a simple mathematical model for describing text and a common way to extract text features [17]. This model disregards the grammatical and sequential aspects of the text, treating it as a collection of words. Each word occurrence in the document is considered as an independent entity.

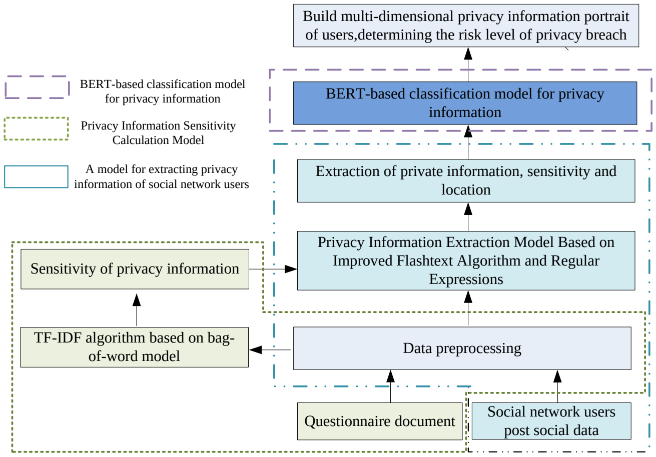

Build multi-dimensional privacy information portrait of users,determining the risk level of privacy breach
BERT-based classification model for privacy information
BERT-based classification model for privacy information
Privacy Information Sensitivity Calculation Model
A model for extracting privacy information of social network users
Extraction of private information, sensitivity and location
Privacy Information Extraction Model Based on Improved Flashtext Algorithm and Regular Expressions
Sensitivity of privacy information
TF-IDF algorithm based on bag-of-word model
Data preprocessing
Questionnaire document
Social network users post social data

Fig. 1. Overall structure of the multidimensional privacy portrait model of social network users.

The TF-IDF algorithm takes into account both the frequency of a term in a specific document and the importance of the term in the entire document collection, thereby reflecting the significance of the term more accurately in the text. By computing the inverse document frequency, the TF-IDF algorithm can consider the distribution of privacy information across the entire corpus and capture the global characteristics of terms. This enables it to calculate privacy sensitivity more accurately and better represent the importance of privacy information to users. Therefore, this paper utilizes the TF-IDF algorithm to calculate the sensitivity of privacy information.

The model for calculating privacy information sensitivity consists of two main modules: Privacy information feature extraction, which involves extracting features from pre-processed text using the bag-of-words model. Privacy information sensitivity calculation, which includes transforming the text into a bag-of-words vector and ultimately calculating the sensitivity of privacy information using the TF-IDF algorithm.

1) Privacy information feature extraction: The structure diagram of privacy information feature extraction is shown in Fig. 2.

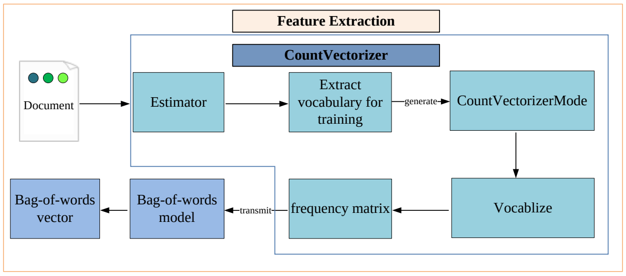


Fig. 2. Privacy information feature extraction structure diagram.

The specific steps are.

Step 1: Construct a corpus Document (pre-processed survey questionnaire document).

Step 2: Training a CountVectorizer model and creating CountVectorizerMode, extracting privacy information words in the document for training and obtaining the vocabulary i.e. all privacy information words present in the corpus.

Step 3: Transforming the text into a word frequency matrix of documents regarding privacy information words.

Step 4: Build a word bag model and convert the word frequency matrix into a word bag vector.

2) Privacy information sensitivity calculation: In this stage, the TF-IDF (sensitivity w_i) value of privacy information is calculated based on the bag-of-words vector. The specific steps are.

Step 1: Calculate the TF value. The TF value depends only on the number of times the privacy information words are in the corpus and is calculated as:

$$TF = \frac{D_{w,\mathrm{d}}}{\sum_k D_{k,d}} \quad (1)$$

where, D denotes the corpus, the symbol $D_{w,d}$ denotes the number of occurrences of privacy word w in document *d*, and the denominator denotes the total number of occurrences of all privacy words w in D.

Step 2: Calculate the IDF value, which is calculated as:

$$IDF = \log \frac{|D|}{DF_{w,D} + 1} \quad (2)$$

where, |*D*| denotes the total amount in the corpus, $DF_{w,D}$ denotes the number of documents containing the privacy word w. To avoid the denominator and the occurrence of arithmetic errors, the IDF needs to be optimized, i.e., denominator + 1.

Step 3: Calculate TF-IDF value.TF-IDF value of a privacy word is the product of TF and IDF, which is calculated as:

$$TF - IDF = TF * IDF \quad (3)$$

B. *A Model for Extracting Privacy Information of Social Network Users*

The FlashText algorithm, created by Singh V, is designed to specifically match complete English words. It serves as a novel and efficient algorithm for keyword search and replacement [18]. However, considering the disparities between English and Chinese texts, this paper enhances the FlashText algorithm to cater to Chinese texts. The enhanced algorithm, known as FlashText+, incorporates jieba word segmentation and constructs a Chinese Trie dictionary. By doing so, the FlashText+ algorithm not only retains the advantages of the original FlashText algorithm but also becomes applicable in Chinese scenarios. Fig. 3 illustrates the structural diagram of the model for extracting user privacy information in social networks.

In the privacy information extraction stage, in order to facilitate the extraction of user privacy information, this paper have constructed two Chinese privacy information dictionaries, dic1 and dic2, which store two parameters, respectively, where dic1 = (privacy word, sensitivity) and dic2 = (keyword, sensitivity). The privacy word indicates user privacy information that can be accurately collected and defined, while the keyword indicates user privacy information that contains a specific keyword or word.

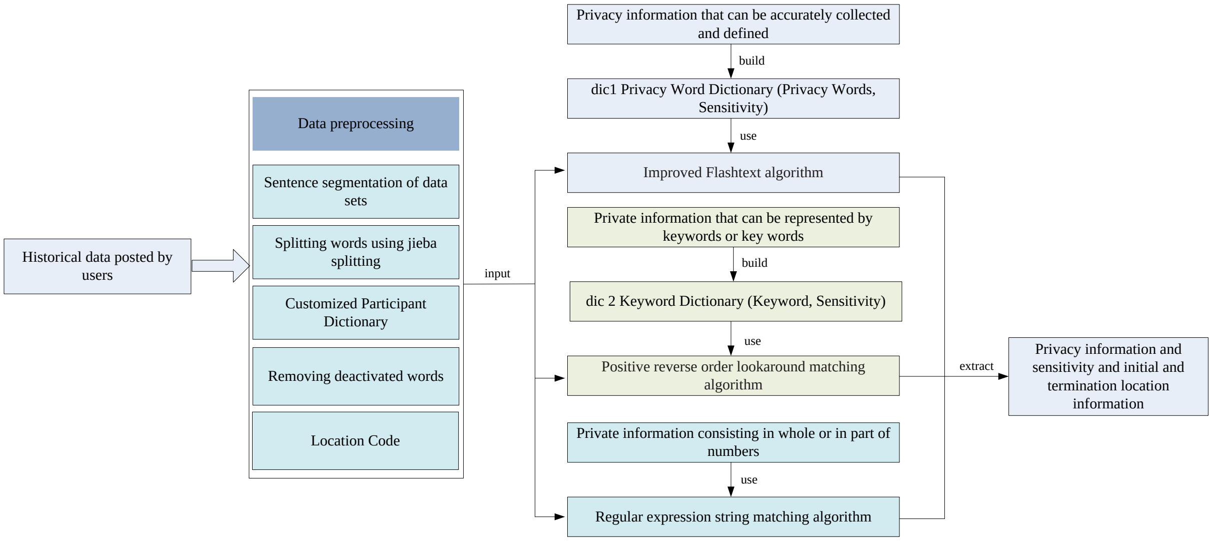

Historical data posted by users
Data preprocessing
Sentence segmentation of data sets
Splitting words using jieba splitting
Customized Participant Dictionary
Removing deactivated words
Location Code
input
Privacy information that can be accurately collected and defined
build
dic1 Privacy Word Dictionary (Privacy Words, Sensitivity)
use
Improved Flashtext algorithm
Private information that can be represented by keywords or key words
build
dic 2 Keyword Dictionary (Keyword, Sensitivity)
use
Positive reverse order lookaround matching algorithm
Private information consisting in whole or in part of numbers
use
Regular expression string matching algorithm
extract
Privacy information and sensitivity and initial and termination location information

Fig. 3. Structure of social network user privacy information extraction model.

In this paper, this study classify users' privacy information into three types, i.e., privacy information that can be accurately collected and defined, privacy information that cannot be accurately collected and defined but can be represented by a keyword or word, and privacy information that consists entirely or partially of numbers. The first type of privacy information extracted using the improved FlashText algorithm; for the second type of privacy information, the keywords defined in dic2 are used to extract the user's privacy information and its sensitivity and its location using the PROLM algorithm; for the third type of privacy information, the corresponding rules are defined to extract the privacy information, sensitivity and its location using the string matching algorithm.

Example 1. A message posted by a user T = {Punch up Henan University, near Longting District.}

1) Extracting user privacy information using an improved FlashText algorithm.

The improved FlashText algorithm extracts user privacy, sensitivity, and location information in the following steps.

Step 1: Data pre-processing. Data preprocessing is performed on the user-posted information T. The word order and location information of T after preprocessing are shown in Fig. 4.

Step 2: Build a Chinese Tree dictionary.

Construct a Chinese tree dictionary using the privacy words in dic1 as input. First, create an empty node root which is the starting point of all private words, then insert all private words into the Trie tree one by one, if there is already a node pointing to the current character in the path during the insertion process, then visit the node, if there is no corresponding node, then create a new node pointing to the current character, if a private word completes the insertion, mark the last node in its path. The process is as in Algorithm 1.

Algorithm 1: Build a Chinese Tree dictionary

```
Input: P= {$p_1$, $p_2$, ..., $p_r$}
Output: Chinese Tree dictionary
Create an initial state empty root 0
For i∈1, 2 ,....., r Do
    Current ← root
    j ← 1
    Pi ← Current ($P_i$)
    While j ⩽ $m_i$ AND next (Current, $P_i^{\,j}$) ≠ end Do
        Current ← next (Current, $P_i^{\,j}$)
        j←j+1
    End of While
    If j ⩾ $m_i$ Do
        |Create a new non empty node Stat
            next (Current, $P_i^j$ ) ←State
            Current ← State
            j← j+1
            break
    End If
    If Current is already terminal Then F(Current)←F(Current)∪{i}

    Else mark Current as end, F(Current)←{i}
End of For
```

This study defines some of the notations used in Algorithm 1. *P* is the privacy word in dic1, root is the root node, *i* is the position of the privacy word, *j* is the position of the node in the tree dictionary, *Current (P_i)* is the position of the node where the current keyword is located, *next (Current, P_i^j)* is the position of the node where the next keyword is located, and *F* is the transfer function.

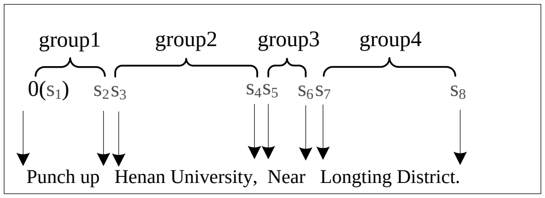


Fig. 4. Phrase order and position information of L after pre-processing.

Step 3: Iterate through the Chinese tree dictionary and g output the results. The pre-processed T is used as input, and each node in the Chinese tree dictionary is traversed one by one in terms of phrases, starting from the root node root (0) for matching. If the characters match, the current state jumps to the corresponding state. When the output state (the state with shaded background) is reached, the corresponding private word for the sequence of matched nodes is output, and the sensitivity is returned. Additionally, the location of the privacy information is saved. If the characters do not match, the next matching step is performed. The traversal process is shown in Algorithm 2.

Algorithm 2: Privacy Information Extraction

```
Input: P={$p_1$,$p_2$,...,$p_r$}, T=($t_1$,$t_2$,...,$t_n$)
Output: Privacy Words pi, Sensitivity, Start and end positions
Preprocessing
        FT← Build_FT(P)
Searching
    Current ← Initial state of the Flashtext FT
    For pos ∈1,2,...,n Do
            While $next_{FT}$ (Current, $t_{pos}$) = end AND $S_{FT}$(Current) ≠ end Do
                Current ← $S_{FT}$(Current)
            End of While
            If $next_{FT}$ (Current, $t_{pos}$) ≠ end Then
                Current ← $next_{FT}$ (Current, $t_{pos}$)
            Else Current ← initial state of FT
            End of If
            If Current is terminal Then
                Mark all the occurrences (F(Current), pos)
            End of If
    End of For
F
```

This study define some symbols used in Algorithm 2, P is the privacy word in dic1, T is the input text, *$next_{FT}$* is the next matching node position, *S_{FT}* is the current matching node position, and *F* is the transfer function.

2) *Extracting* user privacy information using the PROLM algorithm.

PROLM algorithm focuses on matching results and omits duplicated content, enhancing matching speed by searching from right to left. However, it cannot match privacy information consisting of digits alone. A combination with a string matching algorithm is proposed to extract such information.

PROLM algorithm：Matching text from right to left can be abstractly represented as: (?<=SubExp1)|(?<=SubExp2).

According to the characteristics of SubExp1 can be summarized into three categories.

- The subexpression SubExp1 is of fixed length and is in general mode.
- The subexpression SubExp1 has a variable length, for non-greedy mode.
- The subexpression SubExp1 is not fixed in length, but it contains match-first quantifiers, which is a greedy mode.

The matching principle of the PROLM algorithm is shown in Fig. 5, which can be divided into main matching and sub-matching.

Its main idea is to extract phrases containing keywords in dic2 from data published by users.

Taking the text in Fig. 5 as an example, the main matching process is shown Algorithm 3.

Algorithm 3: Main matching

```
Input: T={t_1,t_2,...,t_n}, dic2={k_1， k_2,....,k_n}
Output: Privacy Words, Sensitivity, Start and end positions
for i∈1,2,...,n Do
    i← 1
    while there is no keyword ki in dic2 in T Do
    end of while
    if T contains the keyword ki in dic2 Do
        find the keyword position Si from the initial position Sk
        to the right
        S_k←S_{i-len}(SubExp1)
        Enter sub matching
    End of If
    if next round of sub matching is required
        give control to subsequent word expressions
    else location of report Si, matching failed
    end of if
    if Subsequent sub expression matching succeeded
        location of report Si , match succeeded
    else location of report S_i, matching failed
    end of if
End of for
```

This study define some of the symbols used in Algorithm 3, T is the input text, k_n is the keyword in dic2, S_k and S_i are the location of the text, and *len(SubExp1)* is the length of "SubExp1".

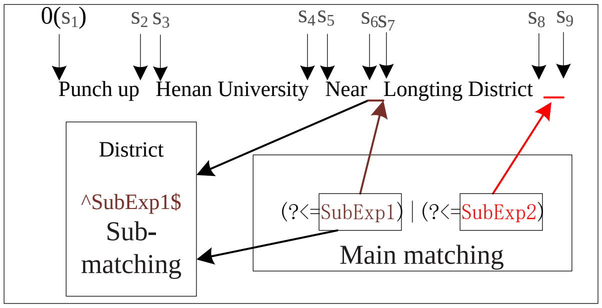


Fig. 5. PROLM algorithm matching principle.

The main matching process is divided into the following steps.

Step 1: Matching is attempted from position 0 to the right, before finding the position that satisfies (? <=SubExp1) and contains the keyword in dic2, the matching must fail until position S_9 is found and its requirements are met.

Step 2: Locate the position S_6, which satisfies the minimum length requirement for "SubExp1", by moving leftwards from S_9.

Step 3: Perform a sub-matching process by applying "SubExp1" from position S_6 and moving rightwards.

Step 4: Successfully match "(? <=SubExp1)" and proceed to the subsequent sub-expression "(? <=SubExp2)". Keep attempting to match until the entire expression either matches or fails. Report whether the entire expression at position S9 matches successfully or fails.

Step 5: If necessary, continue to find the next position S_5 and start a new round of attempted matching.

The sub-matching process is divided into the following main steps.

Step 1: Upon entering sub-matching, the source string has been determined as the string between S9 and S6, and the regular expression at this point becomes "^SubExp1$". In this round of sub-matching, once a match is successful, it must start at S9 and end at S6.

Step 2: Once the sub-expression is fixed, whether the match succeeds or fails, the match result is returned, and there is no need for further rounds of matching.

Step 3: When the length of the subexpression is not fixed, in the case of a greedy mode, if the match fails, it is reported as a failure and the next round of matching is requested. If the match is successful, all backtracking states are discarded and the success is reported, eliminating the need to try the next round of matching.

Step 4: In the case of a greedy mode, if the match fails, it is reported as a failure and the next round of matching is requested. If the match succeeds, all backtracking states are discarded, the success is reported, and the successful content of this match is recorded. The next round of matching is requested until the longest match is obtained.

The sub-matching process is shown in Algorithm 4.

Algorithm 4: Sub matching

```
Input: $S_k$, $S_i$
Output: return to main matching
Sub Match Start
    if fixed length of sub-expressions do
        report matching results, no need for next round of sub matching
    else
        next
    en of if
    if subexpression is not greedy mode and matching succeeded
        report matching results, no need for next round of sub matching
    else
        matching failure, request to enter the next round of sub matching and return to main matching
    end of if
    if the sub expression is greedy and matched successfully
        matching success, record this success, request to enter the next round of sub matching
    end of if
    if the sub expression is greedy and matched failure
        return to main matching
    end of if
```

3) Extracting user privacy information using regular expression string matching algorithm.

Use regular expression string matching algorithm to define corresponding matching rules to extract privacy information and sensitivity from the preprocessed text for privacy information that is entirely or partially composed of numbers, such as a mobile phone number, license plate number, mailbox, etc.

C. BERT-Softmax Classification Model for Privacy Information

1) BERT model: BERT is a pre-trained language model based on a Transformer encoder, which enables reading the whole text at once for bi-directional linguistic representation to achieve the prediction task, and the input Embedding is encoded and transformed through layers of Encoder. In this paper, a pre-trained BERT-based model is utilized to classify the extracted user-privacy information.

Compared to traditional text classification methods, text classification based on BERT pre-trained models has the following advantages:

- Better semantic understanding: The BERT model can learn contextual information from the text, leading to a better understanding of the text's meaning and improving the accuracy of text classification.
- Improved robustness: The BERT model can handle texts of different lengths, enhancing the model's robustness and enabling it to process various types of text data.
- Good generalization capability: Text classification models based on BERT can be fine-tuned to adapt to different text classification tasks, thereby exhibiting good generalization capability.

The structure of the BERT model [19] is shown in Fig. 6. It consists of a multilayer bidirectional Transformer encoder. Where, the middle Trm denotes the Transformer encoder, E_1, E_2,...E_N denotes the text input vector of the word, the vectorized representation of the text is obtained after the Trm module, and T_1, T_2,...T_N denotes the final text representation.

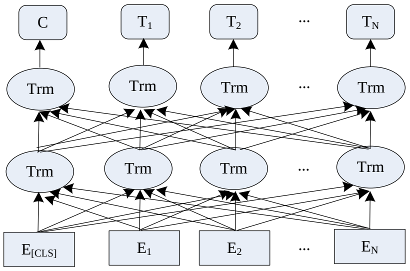


Fig. 6. Structure of the BERT model.

2) BERT-Softmax classification model. The structure of the BERT- Softmax classification model is shown in Fig. 7.

The classification process is as follows: the extracted user-privacy information is used as input text data, the text features are extracted using the BERT pre-training model, i.e., the text is encoded using one-hot, the unstructured text sequence is converted into a structured feature vector, the text features are input into the fully connected layer and the probability of each type of label is calculated by the Softmax function, and the label corresponding to the maximum probability is the result of the model classification.

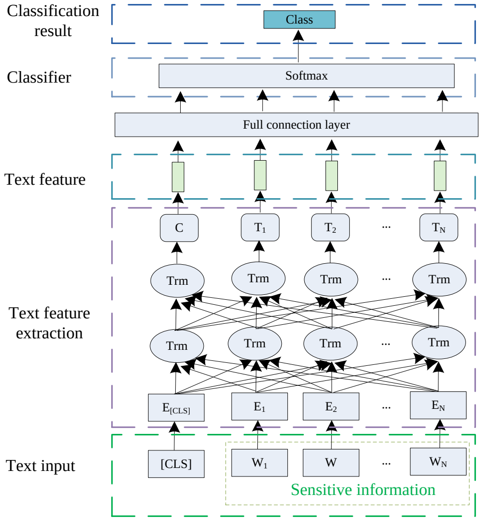


Fig. 7. Structure of BERT-based pre-trained classification model.

3) Attention mechanism: whose core idea is to calculate the interrelationship between each word in the input text and all the words in that text, and to measure the relevance and importance of different words in the input text, and to adjust the weight of each word by these interrelationships to obtain a new expression for each word, which contains not only the semantics of the word itself but also its relationship with other words [20].

The calculation process of Self-attention for a single word is shown in Fig. 8, where Query denotes the current word to go to a query concerning other words, Key denotes waiting to be checked, and Value denotes the actual feature information.

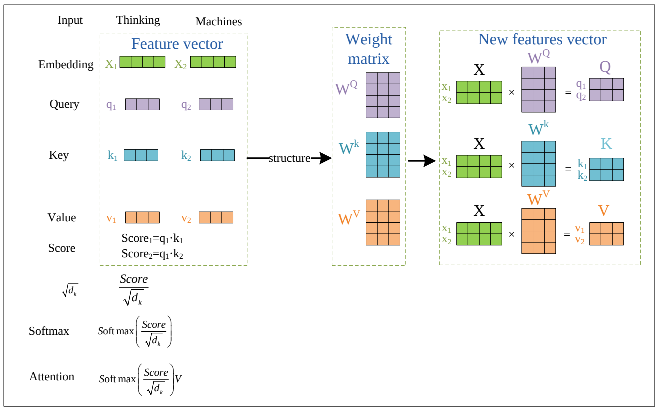


Fig. 8. Self-attention calculation process for a single word.

The calculation steps are as follows:

Step 1: For each word of the input after one-hot encoding get the initial feature vector noted as X=[x_1,x_2], and initialize three weight vectors q, k, v;

Step 2: With training the entire weight matrix W^Q, W^K, W^V is constructed;

Step 3: Use Eq. (4) to obtain three new feature expression matrices Q(Query), K(Key), and V(Value) for each word;

$$\begin{cases} Q=XW^Q \\ K=XW^K \\ V=XW^V \end{cases} \tag{4}$$

Step 4: Calculate the inner product Score using Eq. (5), which indicates the relationship between the current word and other words.

$$Score = QK^T \tag{5}$$

Step 5: To prevent the Score from increasing with vector dimension, the inner product is calculated by dividing by the factor $\sqrt{d_k}$, d_k being the vector dimension of the input information and normalized to a probability distribution by the softmax function.

Step 6: The distribution of scoring by inner product is the weighted average of Value, and Attention is calculated using Eq. (6).

$$Attention(Q,K,V) = \mathrm{Soft\,max}\left(\frac{Score}{\sqrt{d_k}}\right)V \tag{6}$$

In practice, the Transformer encoder uses Multi-head attention mechanism, Multi-head mechanism that is Multi-head attention is a network optimization technique used in BERT network structure, using different heads to focus on different context dependency patterns, similar to the model integration effect, which can achieve parallel operations, the input of the network into multiple branches, respectively, do attention mechanism, and finally, the results of each branch will be spliced to get, its calculation as in Eq. (7) and Eq. (8).

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{7}$$

$$Multi-head(Q,K,V) = Concat(head_i)W^O \tag{8}$$

where, W_i^Q, W_i^K, and W_i^V represent the W^Q, W^K, and W^V weight matrices of the ith head, W^O denotes the mapping vector of Multi-head, and Concat(·)denotes the splicing function.

4) Classification: In this paper, introduce softmax regression model for privacy information classification. The softmax regression, like linear regression, does a linear superposition of the input features and ourights. One major difference from linear regression is that the number of output values in softmax regression is equal to the number of categories in the label. Assuming there is a training sample set {(x^1, y^1), (x^2, y^2),..., (x^m, y^m)}, where x^i represents the privacy information vector corresponding to the i-th training sample, with a dimension of n for a total of m training samples, $y^i \in$ (1,2,..., n) represents the category corresponding to the i-th training sample, and n is the number of categories. For a test input sample x, the distribution function of the Softmax regression model is the conditional probability p(y=j|x), indicating the probability that x belongs to category j, where the category with the highest probability of occurrence is the category to which the current sample belongs, and the hypothesis function of belonging to each category is as in Eq. (9).

$$\mathrm{h}_\theta(x^i) = \begin{bmatrix} p(y^i=1|x^i;\theta) \\ p(y^i=2|x^i;\theta) \\ \dots \\ p(y^i=n|x^i;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{n} e^{\theta_j^T x^i}} \begin{bmatrix} e^{\theta_1^T x^i} \\ e^{\theta_2^T x^i} \\ \dots \\ e^{\theta_n^T x^i} \end{bmatrix} \tag{9}$$

Among them, the probability that any element *$p(y^i=n|x^i;\ \theta)$* in *$h_\theta(x^i)$* is the current input sample x^i belonging to the current category *n*.

For a data set with *m* training samples, the cost function of the Softmax regression model is as in Eq. (10).

$$J_\theta = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} I\left\{y^{(i)}=j\right\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{n} e^{\theta_j^T x^{(i)}}}\right] \tag{10}$$

where, *I{.}* is the indicative function, *m* is the number of samples, *n* is the number of categories, *i* denotes a certain sample, x^i denotes the vector representation of the *ith* sample *x*, and *j* denotes a certain category.

After obtaining the model parameters *θ*, the probability of belonging to category *j* for the sample *x* to be measured is:

$$\mathrm{p}\left(y^{i}=j \mid x^{i};\theta\right)=\frac{e_{j}^{T}x^{i}}{\sum_{l=1}^{n}e_{l}^{T}x^{i}} \qquad (11)$$

The probability of x belonging to all categories is calculated using Eq. (11), and the category with the highest probability is chosen as the classification category for x.

D. Building a Portrait of User Privacy Information

The construction of user privacy portrait first extracts user privacy information through the PROLM_ FlashText+_ SMA algorithm, and then uses BERT-Softmax to classify the privacy information. Then, use Eq. (12) to calculate the average sensitivity of privacy information to determine the level of user privacy leakage risk. Finally, use word cloud technology to construct user portrait of privacy information.

$$W=\frac{\sum_{\mathrm{i}=0}^{n}w_{i}}{\mathrm{n}},(n\geq 0,i\geq 0) \qquad (12)$$

where, W denotes the average sensitivity of privacy information, n denotes the number of privacy information, i denotes the i-th privacy word, and w_i denotes the sensitivity of the i-th privacy word.

Set a threshold to divide the risk of user privacy leakage into three levels.

Heavy leakage risk: average sensitivity $W > 0.11$.

Moderate leakage risk: average sensitivity $0.07 < W < 0.11$.

Mild leakage risk: average sensitivity $W < 0.07$.

IV. Experimental Results and Analysis

Due to the lack of public datasets related to the precise expression of user privacy, the experimental datasets for user privacy information in this paper are collected by crawlers and collected manually. The dataset in this paper is divided into four parts: 1071 valid questionnaire data for user privacy information sensitivity calculation, 1000 social network users' microblogging history data, a dictionary containing 7673 user privacy words and a dictionary containing 199 user privacy keywords for privacy information extraction, and 60,000 user privacy data for privacy information classification. Experiment in this paper is divided into three parts: user privacy information sensitivity calculation, user privacy information extraction, and user privacy information classification.

A. User Privacy Information Sensitivity Calculation

The dataset used in this experiment is a 1071 valid survey questionnaire, which defines a total of 62 types of privacy information.

This paper uses the BOW-TF-TDF algorithm to calculate privacy sensitivity, and arranges the sensitivity calculation results in descending order, as shown in Table II.

TABLE II. Sensitivity of User Privacy Information

Privacy words	Sensitivity	Privacy words	Sensitivity
ID number	0.155	Purchase preferences	0.125
Email	0.154	Private financial list	0.124
health condition	0.154	Payment records	0.124
phone number	0.153	Income situation	0.123
Home Address	0.153	Purchase Record	0.123
Property status	0.152	major	0.122
Bank card number	0.152	Purchase frequency	0.121
Card Number	0.152	Consumer credit	0.121
academy	0.152	Shopping habits	0.121
Political landscape	0.151	Consumption amount	0.12
height	0.15	Dressing hobbies	0.118
Current address	0.15	Literary Hobbies	0.118
height	0.15	Sports hobbies	0.116
Gender	0.149	Consumption level	0.116
marital status	0.149	Travel Hobbies	0.115
Age	0.143	Game Hobbies	0.1118
weight	0.14	Weibo account	0.106
position	0.137	Chat records	0.104
educational	0.137	QQ number	0.102
Online behavior	0.134	Tiktok	0.102
Mode of travel	0.134	Kuaishou	0.102
Life experience	0.134	Station B	0.101
Social relationship	0.133	WeChat signal	0.099
Family member	0.133	Alipay account	0.096
Device Information	0.133	The Little Red Book	0.095
Work unit	0.132	Investment hobbies	0.013
Personal itinerary	0.132	Art Hobbies	0.011
IP address	0.132	Pets	0.003
name	0.13	constellation	0.0001
occupation	0.13	religious belief	0.00002
Training experience	0.13	blood group	0.00002

According to Table II this paper classifies privacy information into three levels.

High privacy: when wi >0.13, the privacy information includes the user's ID number, home address, name, IP address, and other information, which can accurately locate the user's identity and location once exposed.

Moderate privacy: when 0.13 >wi >0.1, the privacy information includes information about the user's hobbies, social behavior habits, etc., which can only be roughly understood about the user even if exposed.

Mild privacy: when $w_i < 0.1$, this type of information, even if exposed, will not have a major impact on the user.

B. User Privacy Information Extraction

The dataset used in this experiment is the basic Weibo information of 1000 social network users and 65536 historical and dynamic information. This paper evaluates the superiority of this scheme based on the precision of privacy information extraction.

1) Data preprocessing: This paper mainly collects data from Weibo users on social networking platforms.

Data preprocessing mainly includes filtering the empty data in the dataset, using the Jieba word segmentation library to segment the dataset based on the stop list proposed by the Harbin Institute of Technology laboratory, and removing stop words, special symbols, and meaningless words.

2) Evaluation indicators: This paper evaluates the performance of different models by extracting privacy information precision.

The goal of privacy information extraction is to correctly extract all the privacy information from the historical data posted by users, and Precision is an important indicator of how good the information extraction model is. Precision is calculated by the following formula.

$$\Pr ecision = \frac{|\,\mathrm{privacy}_u\,|}{|\,extreact_u\,|} * 100\% \qquad (13)$$

where, $|privacy_u|$ denotes the number of extracted privacy information, $|extreact_u|$ denotes the total number of extracted information.

This paper selects classic text information extraction models BILSTM, BILSTM-CRF, and LSTM-CRF as comparative experiments. We divided the dataset into training, validation, and testing sets in a ratio of 8:1:1. The distribution of the dataset is shown in Table III.

TABLE III. PRECISION COMPARISON OF DIFFERENT PRIVACY INFORMATION EXTRACTION MODELS

Model	Precision
BILSTM	87.92
BILSTM-CRF	90.32
LSTM-CRF	89.64
PROLM_FlashText+_SMA	**93.63**

According to the data in Table III, it can be seen that the PROLM_FlashText+_SMA algorithm proposed in this paper has the highest precision rate in privacy information extraction tasks (93.63), followed by BILSTM-CRF (90.32), and the model BILSTM has the lowest precision rate (87.92). Experiments have shown that the PROLM_FlashText+_SMA algorithm outperforms the comparative model in terms of precision in extracting privacy information.

C. User Privacy Information Classification

1) Experimental dataset: The classification dataset used in this paper is the THUCNews text classification dataset provided by the Tsinghua NLP group and the author collection of 60,000 pieces of data about users' privacy information, containing three types of data: highly privacy, moderately privacy and mildly privacy information, and this paper divide the dataset into the training set, validation set, and test set according to 8:1:1, and the distribution of privacy information classification dataset is shown in Table IV.

TABLE IV. CLASSIFICATION OF PRIVACY INFORMATION CLASSIFICATION EXPERIMENTAL DATASET

Category	Label	Train Set	Test Set	Validation Set
Highly privacy	1	8000	1000	1000
Moderately privacy	2	8000	1000	1000
Mildly privacy	3	8000	1000	1000

2) Model and parameter setting: The experimental environment is Windows 10, based on the PyTorch framework, CUDA version 11.7.1, CUDANN 8.5, and the GPU used to accelerate the training is RTX3050, and the model parameters are set as shown in Table V.

TABLE V. PARAMETERS OF THE MODEL

Parameters	Numerical value
hidden_size	768
epoch	3
batch_size	32
pad_size	32
learning_rate	5e-5
Attention Dimension	64

3) Evaluation indicators: The common evaluation metrics for classification problems are the Precision, Recall, and F1 (F1-Score) index. Precision and recall are important metrics to measure how good an information extraction model is. The F1-score is the summed average of Precision and Recall, which is used to evaluate the precision and recall together.

The precision calculation formula equation is as follows.

$$Precision = \frac{\mathrm{Sum}_{true}}{Sum_{forecast}} \qquad (14)$$

where, Sum_{true} indicates the number of privacy information correctly classified, and $Sum_{forecast}$ indicates the number of privacy information predicted to be in that category.

The recall is calculated as follows.

$$Recall = \frac{Sum_{true}}{Sum_{actual}} \qquad (15)$$

where, $Sum_{actural}$ indicates the actual amount of privacy information in that category.

The formula for calculating the F1-Score is as follows.

$$F1 = \frac{\Pr ecision * \mathrm{Re}\, call}{\Pr ecision + \mathrm{Re}\, call} * 2 \qquad (16)$$

In this paper, we choose the classical text classification models BERT_CNN, BERT_RNN, BERT_RCNN, and ERNIE as the comparison experiments. The training set, validation set, and test set data are kept consistent with the BERT model during the training process, and the Precision, Recall, and F1 index of different models are compared as shown in Table VI.

TABLE VI. COMPARISON OF PRECISION, RECALL, AND F1 INDEXES OF DIFFERENT CLASSIFICATION MODELS

Model Name	**Precision**	**Recall**	**F1-Score(F1)**
BERT_CNN	0.9231	0.9530	0.9378
BERT_RNN	0.9382	0.9550	0.9419
BERT_RCNN	0.9384	0.9640	0.9510
ERNIE	0.9412	0.9730	0.9568
BERT-Softmax	**0.9846**	**0.9750**	**0.9798**

From Table VI, it can be seen that the privacy information classification model proposed in this paper based on BERT Softmax has an accuracy and recall rate of over 0.97 in privacy information classification experiments, and the F1 index reaches 0.9798. From the overall experimental results, it can be seen that the F1 value of the BERT model is higher than that of the BERT_CNN model is four percentage points higher than the ERNIE model by two percentage points, indicating that this model is compared to BERT_CNN, BERT_RNN, BERT_RCNN and ERNIE classification models can better represent the semantic information of short text privacy information and have better privacy information classification performance.

V. CONCLUSION

In order to address the issue of user privacy and social relationships being prone to privacy leakage during data publishing and information sharing, this paper proposes a user privacy information portrait model for social networks. To address the issue of difficulty in measuring user privacy information sensitivity, the BOW-TF-IDF algorithm is used to calculate user privacy information sensitivity; Due to the diversity of user privacy information, this paper proposes the PROLM_FlashText+_SMA privacy information extraction algorithm, which extracts the privacy information exposed by users from historical data published on the social Internet of Things. The privacy information is classified into high, moderate, and mild using the BERT-Softmax based privacy information classification model. Finally, Use the extracted privacy information to construct a multidimensional user portrait of privacy information and determine the level of privacy leakage risk by calculating the average sensitivity of privacy information. The experiment shows that the privacy information extraction model proposed in this paper exhibits its superiority in terms of accuracy; the privacy information classification model proposed in this paper outperforms traditional classification models in terms of accuracy, recall, and F1 index. It can better represent the semantic information of short text privacy information and has better privacy information classification performance.

In the future, this study will be combined with trust between users to generate access controls to protect the privacy of users in social networks.

REFERENCES

[1] T. L. Gu, F. R. Hao, L. Li, J. J. Li and L. Chang, “Behavior Accountability of Agents Responsible for Privacy Negotiation in Social Networks,” Ruan Jian Xue Bao/Journal of Software, vol. 33, no. 9, pp. 3453-3469, 2022.

[2] R. N. Xie, X. N. Fan, Y. Lin et al., “Research on extended access control mechanism in online social network,” Chinese Journal of Network and Information Security, vol. 7, no. 5, pp. 123-131, 2020.

[3] Al-Asbahi R, “Structural Anonymity For Privacy Protection In Social Network,” International Journal of Scientific and Research Publications (IJSRP), vol. 11, no. 6, pp.102-107, 2021.

[4] C. Lian and Z. Chen, “Anonymous privacy protection algorithm based on sensitive attribute classification,” 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), pp. 222-226, 2020.

[5] X. Yin, S. Zhang and H. Xu, “Node Attributed Query Access Algorithm Based on Improved Personalized Differential Privacy Protection in Social Network,” International journal of wireless information networks, vol. 26, no. 3, pp.165-173, 2019.

[6] B. Ning, Y. Sun, X. Sun, and G. Li, “Differential privacy protection on weighted graph in wireless networks,” Ad Hoc Networks, vol. 110, no.102303 pp. 1-10, 2021.

[7] H. Huang, D. Zhang, F. Xiao et al., “Privacy-Preserving Approach PBCN in Social Network With Differential Privacy,” IEEE Transactions on Network and Service Management, vol. 17, no. 2, pp. 931-945, 2020.

[8] Z. Y. He, Q. H. Zhu and M. Bai, “The Construction of Urban Elderly User Portrait from the Perspective of Pension Service,” Journal of Information, vol. 40, no. 9, pp. 154-160, 2021.

[9] X. Li and S. He, “Research and Analysis of Student Portrait Based on Campus Big Data,” 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA), pp. 23-27, 2021.

[10] M. H. You, Y. F. Yin, L. Xie and S. L. Lu, “User profiling based on activity sensing,” Journal of Zhejiang University(Engineering Science ）, vol. 55, no. 4, pp. 608-614, 2021.

[11] Z. Zhang, L. Han, M. Chen, “Fuzzy MLKNN in Credit User Portrait,” Appl. Sci, vol, 12, no.22, pp. 11342, 2022.

[12] S. F. Wu, C. C. Wu and J. Zhu, “Microblog User Dynamic Portrait Generation Based on Interest Transfer,” Information Science, vol. 39, no. 8, pp. 103-111, 2021.

[13] Z. Ding, C. Yan, C. Liu, et al., “Short Text Processing for Analyzing User profiles: A Dynamic Combination,” International Conference on Artificial Neural Networks, vol. 12397, pp. 733-745, 2020.

[14] J. Y. Wu and M. Z. Xu, “Video Personalized Recommendation Based on User Portrait and Video Interest Tags,” Information Scence, vol. 39, no.1 , pp. 128-134, 2021.

[15] L. An, J. Y. Hu and G. Li, “Research on profiles of High-impact Users on Social Media in the Context of Emergencies,” Information and Documentation Services, vol. 41, no. 6, pp. 6-16, 2020.

[16] K. S. Jones, “A statistical interpretation of term specificity and its application in retrieval,” Journal of Documentation, vol. 28, no. 1, pp. 11-21, 1972.

[17] X. G. Hu, X. H. Li, F. Xie and X. D. Wu, “Keyword Extraction Based on Lexical Chains for Chinese News Web Pages,” Pattern Recognition and Artificial Intelligence, vol. 23, no. 1, pp. 45-51, 2010.

[18] V. Singh, “Replace or Retrieve Keywords in Documents at Scale,” https://arxiv.org/pdf/1711.00046v2.pdf, 2017.

[19] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” North American Association for Computational Linguistics (NAACL), 2019.

[20] P. Yang and W. Y. Dong, “Chinese named entity recognition method based on BERT embedding,” Computer Engineering, vol. 46, no. 4, pp. 40-45, 2020.

Smart Fruit Identification and Counting using Machine Vision Approach

Madhura R. Shankarpure[1], Dipti D. Patil[2]
Department of Computer Engineering, STES's Smt. Kashibai Navale College of Engineering, India[1]
Department of Information Technology, MKSSS's Cummins College of Engineering for Women, India[2]

***Abstract*—Estimating fruit yield holds significant importance for farmers as it enables them to make precise resource management decisions for fruit harvesting. The adoption of automated image processing technology not only reduces the human labor required but also enhances the accuracy of ripe fruit estimates. This research delves into the performance of an image processing algorithm designed to count and identify oranges. The study employed a multi-phase approach, starting with the creation of a mask to isolate orange content, followed by the detection of circular shapes within the mask. Lastly, the algorithm filters and counts the identified circles. The outcome of this study revealed that the algorithm demonstrated an impressive success rate of approximately 72.4% in correctly identifying oranges with standard deviation of +/- 12.20.**

***Keywords*—Image processing; fruit; multiphase approach; counting**

I. Introduction

Crop yield estimation is important to farmers so that they can accurately predict and manage the resources that will be required to harvest the crop, and to tell when to best pick the crop. The technique of estimation is laborious, occasionally assisted by the use of hand counters. Even though only a few trees are typically examined per block, many man hours are still needed in the hot, humid area [1]. Although multiple load assessments throughout crop growth are ideal, labor constraints make this impossible. Additionally, the sample process has a limit on how accurately the process can estimate the yield of the entire orchard [2]. To meet the requirements of the expanding population while effectively utilizing the resources at hand, sustainable agriculture is important [3]. Precision agriculture, which is aided by cutting-edge sensing and image processing systems [4], artificial intelligence, and other technologies, can produce it. Early in the 1980s, Precision agriculture was created [5]. With the use of Deep Learning architectures and contemporary machine vision, Precision agriculture has a revolutionary impact on a number of agricultural applications, including crop monitoring, disease diagnosis, and intelligent yield calculation. Among these, intelligent fruit yield assessment is crucial in determining the ultimate choices for fruit management and harvesting [6]. An automated image processing technique will allow farmers to reduce the amount of manual labor that is required, and increase the accuracy of the estimated count of ripe fruits. Hannan et al. [7] presents a machine learning system for recognizing orange fruit that consists of segmentation, region labelling, size filtering, perimeter extraction, and perimeter-based detection. Fruits could be clearly differentiated from the background thanks to the segmentation technique, which is based on a Bayesian discriminating analysis. Segmentation, region labelling and perimeter-based detection are all parts of the fruit recognition method. In order to provide adaptive segmentation under changing outside lighting for the segmentation of the orange fruit, the red color space factor was used for augmentation in [8]. Laplacian pyramid transform and fuzzy logic are two image fusion techniques that are used to increase the effectiveness of fruit detection in comparison to Leemans et al.'s [9] use of solely thermal images. In the example in [10], color cues in a picture can be effectively used to segment flaws in "Jonagold" apples. It has been discovered that texture characteristics provide helpful information for evaluating the quality of fruits for example, classifying the grade of apples after dehydration with an accuracy of 95% [11]. The dual color/shape analysis algorithm and laser range-finder model were used to create the spherical fruit recognition system in [12]. The models for illumination and surface reflectance for use in outdoor color vision are explored in [13], with a focus on how to forecast the color of surfaces in outdoor settings. The fruit detection algorithm was created by the authors of [14] using a combination of several observing approaches for a tree canopy. For color recognition, texture and color features are used. Fruit identification uses an effective blend of color and texture. By using a minimal distance classifier and statistical and co-occurrence data obtained from wavelet. processed sub-bands, recognition is accomplished [15]. To distinguish apples with red and green color, color and texture traits are used [16]. According to Femandez et al. [17], a tomato-harvesting robot's stereoscopic vision system has a detection accuracy of about 85%. The author in [18] uses a local or shape-based analysis to quickly identify the fruit and was able to identify it at particular stages of maturation. In order to characterize the characteristics, a key point extraction technique [19] combined corner points and edge points, but it was unable to represent surfaces or an object as a whole. Lokhande et.al. [20] utilize a machine learning technique to evaluate the raw data and determine the trust characteristic. To identify the object a fuzzy control method can be used [21].

This research aims to modify and test a method developed by Payne, Walsh, Subedi and Jarvis to identify ripe Mangos, to identify oranges [22]. To do this however, the traits that make up an orange to human vision must be understood. Oranges have a few distinct features, the first being its bright orange hue, and the second being a rough texture on the skin. Fortunately, oranges are mostly spherical in shape, which proves to be easier when distinguishing between oranges and other objects on an orange tree.

The structure of this paper is as follows. Section II gives the detail explanation of algorithm to identify and count the orange fruit. After that Section III shows the results of developed algorithm with respect to each step. Then Section IV is about discussion, limitation of developed algorithm and finally in Section V, Conclusion is presented.

II. Material and Methods

The process of orange identification and counting is divided into a two-phase approach. In the first phase, the algorithm's objective is to identify and isolate the portions of the image that pertain to oranges. This phase typically involves segmentation and the creation of a mask or region of interest that encompasses the oranges in the image.

The second phase of the process is designed for the actual counting of oranges. This typically involves analyzing the mask or segmented regions generated in the first phase to determine individual orange objects or "blobs." Each blob represents a distinct orange in the image. The algorithm counts the number of these blobs to provide an estimate of the total number of oranges.

A. Orange Isolation

The first phase of the process, as illustrated in Fig. 1, draws inspiration from the work of A. Payne et al. [22] and involves the adaptation of an algorithm originally proposed by Payne for mango crop detection to be applied to oranges. This initial phase is comprised of five distinct steps, each aimed at the application of masks to the image for the purpose of identifying oranges.

In these steps, the image undergoes a comprehensive analysis within both the RGB and YCbCr color spaces, specifically seeking out regions that exhibit the characteristic coloration of oranges. These analyses result in the generation of color masks that highlight the areas matching the orange color profile in both color spaces. Subsequently, these individual color masks are combined to form a single comprehensive orange level mask, which serves as a vital component for the subsequent phase of the process.

This orange level mask is seamlessly transferred to the second phase, where it is employed to facilitate the counting of circular regions within the image. The counting operation essentially determines the total count of oranges present within the image. This two-phase approach offers a systematic and structured method for robust orange identification and counting within images.

The algorithm for orange identification consists of four distinct steps to create masks, which are then combined to form a final mask aimed at isolating orange components within the image.

Step 1: The initial step involves calculating a Normalized Difference Index (NDI) between the red and green color channels for each pixel. The NDI is computed as (g - r) / (g + r) [26]. Subsequently, a threshold is applied to identify regions with NDI values greater than 0. This step serves to accentuate areas with higher redness levels, while minimizing the influence of foliage.

Step 2: While Payne's algorithm originally employs a 3x3 variance filter for each color channel, the approach is adapted for orange identification. Instead, a 3x3 adaptive mean threshold filter is used. This new filter effectively removes excessive foliage from the image and eliminates irrelevant areas, such as the sky, which are unrelated to the oranges.

Step 3: The image is then transformed into the YCbCr color space, and the Cr and Cb channels are extracted for thresholding. The Cr channel measures the red difference within the image, and a threshold condition of pixel value >= 140 is applied, based on the observed average minimum Cr value across various regions in images containing oranges.

Step 4: The Cb channel is employed to evaluate the blue difference within the image. For oranges, a range of lower to medium values is acceptable, as they combine with the Cr channel to produce an orange color. Therefore, a threshold condition of 30 <= pixel value <= 120 is implemented.

After these four stages, the masks generated in each step are combined using Equation 1 to form a final mask. This final mask selectively highlights the orange components within the image:

$$pixel_{finale} = pixel_{NDI} \ \wedge \ pixel_{mean} \wedge pixel_{cr} \ \wedge \ pixel_{cb} \quad (1)$$

By applying these steps and the combination of masks, the algorithm effectively isolates the orange elements in the image, facilitating the subsequent phase of counting and identification.

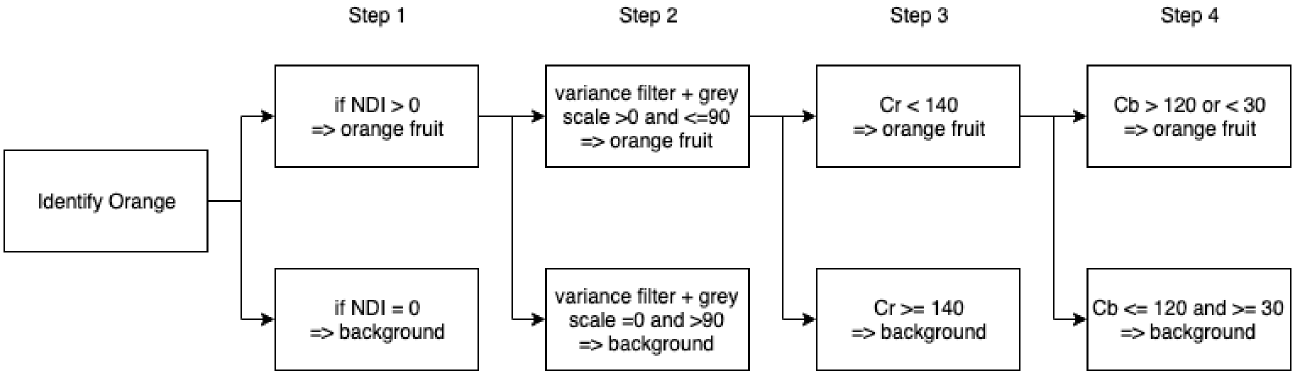


Fig. 1. Outline of the process of the stages of the algorithm.

B. *Orange Counting*

In the second phase, the mask that was generated previously is used to count the number of circles, each circle representing an orange fruit.

The process comprises multiple steps: First, a Gaussian blur is applied to the mask to soften its edges, preparing it for subsequent Canny edge detection. In the second step, a Hough transformation is executed to identify circular regions in the image, with carefully chosen parameter values such as dp, minDist, thresholds, minimum radius, and maximum radius, based on extensive experimentation [23]. In the third step, filtering mechanisms are implemented to refine the detected circles. The filtering involves a rolling average of circle radii, a check for the percentage of pixels within each circle matching the orange criteria established in phase 1, verification that the circle is within the image frame, and ensuring that the circle is not occluded by other detected oranges. These measures are taken to accurately identify oranges among the generated circles.

Algorithm 1: Orange Counting Algorithm

This algorithm is designed to count oranges within a binarized image. It proceeds as follows:

1. Start by applying a median blur to the image to reduce noise.
2. Employ a Hough circle detection technique to identify circular regions within the image. Fine-tune the parameters for best results.
3. Create an empty mask called "found" to keep track of detected circles and initialize the "averageR" variable.
4. For each detected circle:
   - Evaluate the ratio of orange pixels within the circle.
   - Check for overlap with previously detected circles in the "found" mask.
   - Perform certain checks based on circle size and orange pixel ratio to determine confidence in it being an orange.
   - If the circle is highly confident, mark it as a green circle.
   - If moderately confident, mark it as an orange, purple, or red circle based on its size.
   - If not filled enough but the size is close, mark it as maroon.
5. Update the "averageR" based on the newly confirmed oranges.
6. Continue this process for all detected circles.

The result is an image with marked circles in Fig. 2f , with green representing highly confident orange detections and other colors indicating varying levels of confidence. The algorithm filters out overlapping circles to ensure accurate orange counting.

Algorithm 2: Detect Overlapping Circles

This algorithm's purpose is to determine if a circle should be ignored due to excessive overlap with other circles. It operates as follows:

1. Receive input parameters: foundMask, a mask indicating taken (1) and untaken (0) regions, the circle under consideration, the binarized orange image img, and the percentage of orange pixels within the circle (orangePixels).
2. If the percentage of orange pixels within the circle is less than or equal to 50%, return True, indicating that the circle should be ignored.
3. Create a circleMask to represent the area covered by the current circle.
4. Iterate through the image:
   - Check if a pixel falls within the current circle and hasn't been taken (foundMask is False) and the circle mask (circleMask) is True.
   - Keep track of the areas that haven't been taken (areaNotTaken), the area newly covered by the current circle (areaNewOrangeInFrame), and the area that is both not taken and covered by the circle (areaFilledNotTaken).
5. Perform checks to determine if the orange in the circle is distinct and that the non-overlapping section is partly filled. If these conditions are not met, return False.
6. If all conditions are met, return true, indicating that the circle should not be ignored.

This algorithm evaluates circles for potential overlap, ensuring that they are distinct and meet specific criteria before considering them in the orange counting process

Algorithm 3: Count Orange Pixels within a Circle

This algorithm is designed to count the number of pixels within a circular region that have been marked as orange within a binarized image. The steps are as follows:

1. Receive the circle under consideration (\`circle\`) and the binarized orange image (\`img\`) as input, and aim to calculate the percentage of pixels that are marked as orange.
2. Create a \`circleMask\` to represent the area covered by the current circle.
3. Calculate the area of the circle using the formula \`π * circle.radius^2\`.
4. Initialize a counter, \`numOutOfFrame\`, to the area of the circle (i.e., \`circleArea\`).
5. Set the initial count of orange pixels, \`numOrange\`, to 0.
6. Iterate through the height and width of the image.
7. Decrease the value of \`numOutOfFrame\` by 1 for each iteration.
8. Check if the current pixel in the image is both marked as orange (\`img[i, j]\`) and falls within the circle

(`circleMask` is True).

9. If both conditions are met, increment the `numOrange` count by 1.
10. After processing all pixels, check that a sufficient portion of the circle is within the frame. If the percentage of the circle outside the frame exceeds 40%, increase `numOrange` by the number of pixels outside the frame (`numOutOfFrame`).
11. Calculate the percentage of the circle filled with orange pixels within the frame (`percentFilledInFrame`) as a ratio of `numOrange` to the total number of pixels in the image.
12. Calculate the overall percentage of the circle filled with orange pixels, considering both inside and outside the frame (`percentFilled`) as a ratio of `numOrange` to the total number of pixels minus those outside the frame (`numTotal - numOutOfFrame`).
13. Return the minimum of `percentFilledInFrame` and `percentFilled`.

This algorithm assesses the number of orange pixels within a circular region and accounts for pixels both within and outside the frame. The result represents the percentage of orange-filled area within the circle, ensuring a comprehensive measure of orange presence.

III. Results

The algorithm was subjected to a rigorous testing process using a diverse set of online free available orange images that varied in brightness and foliage levels. Algorithm is experimented on set of 10 orange images from [25] which are freely available. Across the entire 10 images, the algorithm exhibited a noteworthy level of success, achieving an approximate 72.4% accuracy in confidently detecting the presence of oranges with standard deviation of +/- 12.20. However, it's important to note that the results were not solely binary, as the algorithm also provided insights into its level of confidence for each image. Fig. 2 shows the correlation of decrease in accuracy of algorithm against actual as fruit count increase. Both trend lines have strong R2 and steady different curves that diverge as fruit count increases.

For images where the algorithm expressed only moderate confidence in its results, typically numbering around one to two images, a specific visual representation was employed to convey this information. In these cases, oranges that the algorithm was highly confident about were distinctly highlighted with green circles. Oranges for which the algorithm expressed moderate confidence were encircled in orange, while those instances where the algorithm had low confidence were marked with either purple or red circles.

Fig. 3 shows step by step results of orange ident and counting also. Fig. 3(a) is original image taken from [24]. Fig. 3(b) is result of the first step in which algorithm is to produce a normalized index between the red and green channels. Fig. 3(c) is step 2 in which an adaptive mean threshold filter replaces the variance filter to improve foliage and sky removal. Fig. 3(d) and Fig. 3(e) shows step 3 and 4 inch which the Cr channel uses a threshold of pixel value >= 140 for medium to high red levels, and the Cb channel employs a threshold condition of 30 <= pixel value <= 120 to detect oranges based on blue differences respectively. In Fig. 3(f) show only orange components of image by applying a mask. Fig. 3(g) and Fig. 3(h) shows the result of orange counting section in which identify the circular regions and filter the circles.

This approach not only provides a quantitative measure of algorithm performance but also offers visual cues to users, making it easier to interpret and potentially refine the results, which can be particularly valuable in scenarios where the algorithm's output may be used for decision-making or further analysis. Fig. 4 shows count of circle for color green, orange, and red, purple in Fig. 3(f). It shows high confidence by detecting green circles as per Algorithm 1.

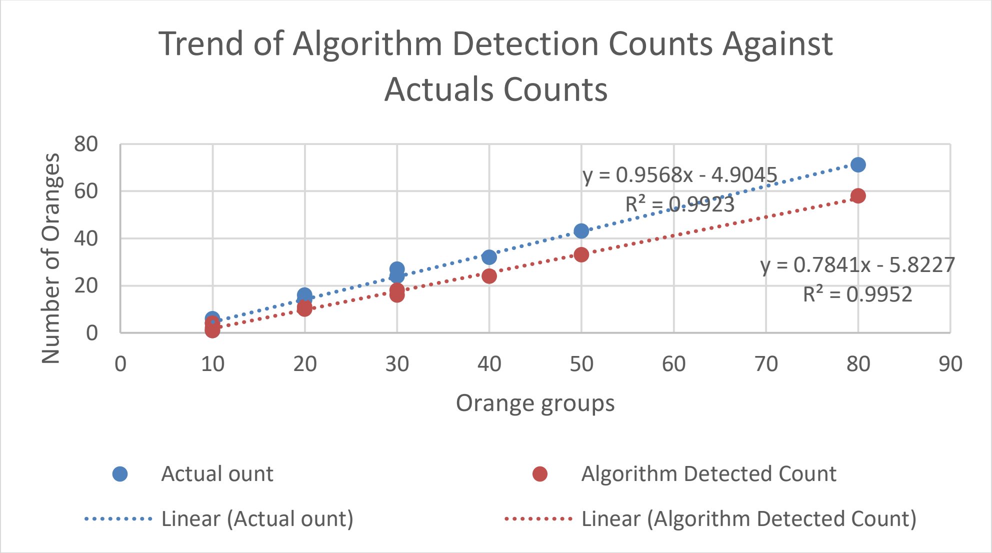
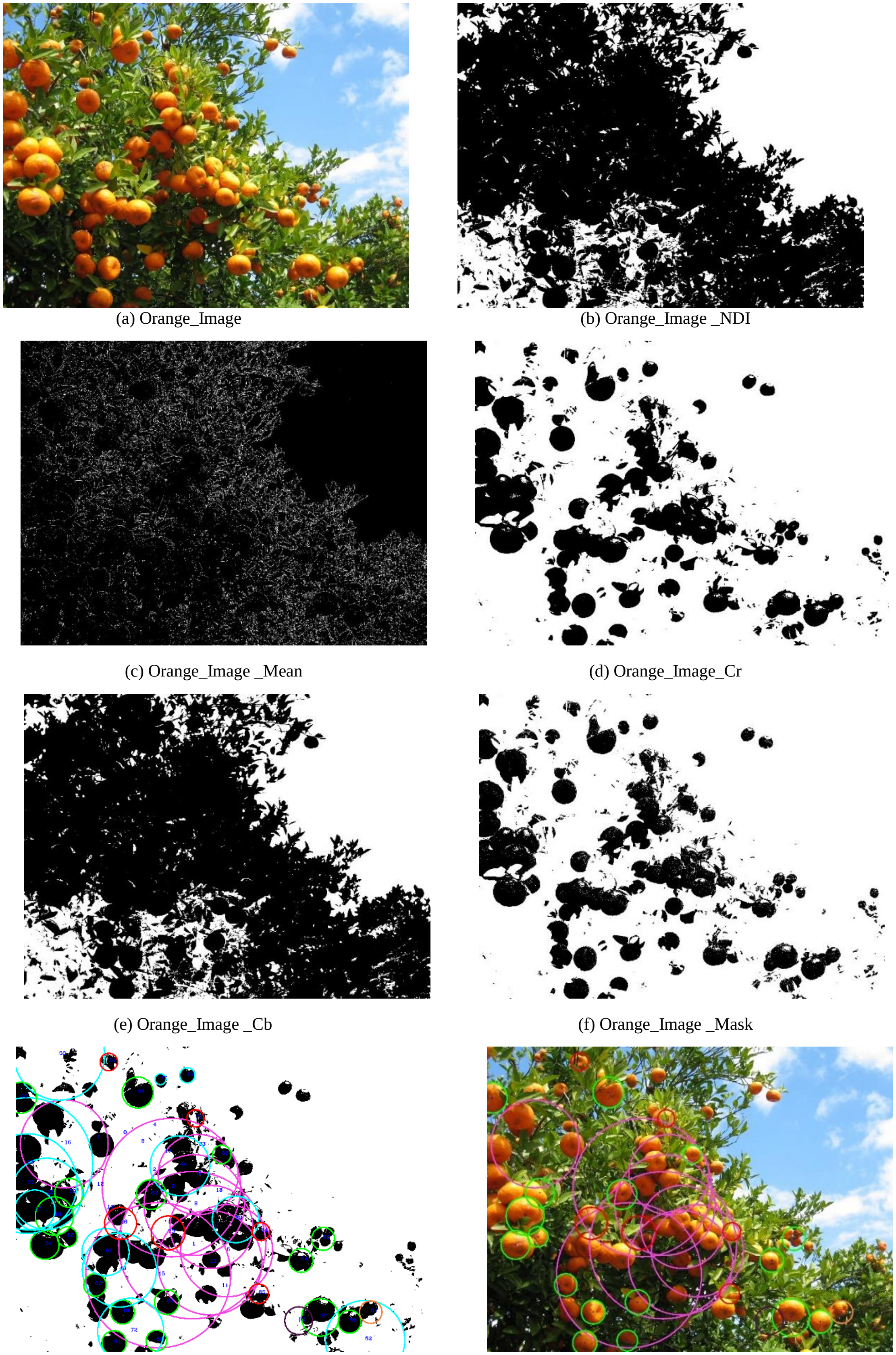


Fig. 2. Trend of algorithm detection counts against actuals counts.



(a) Orange_Image

(b) Orange_Image _NDI

(c) Orange_Image _Mean

(d) Orange_Image_Cr

(e) Orange_Image _Cb

(f) Orange_Image _Mask

(g) Orange_Image _Circle

(h) Finale_Result_Image

Fig. 3. An example image going through each step to detect the oranges within the image.

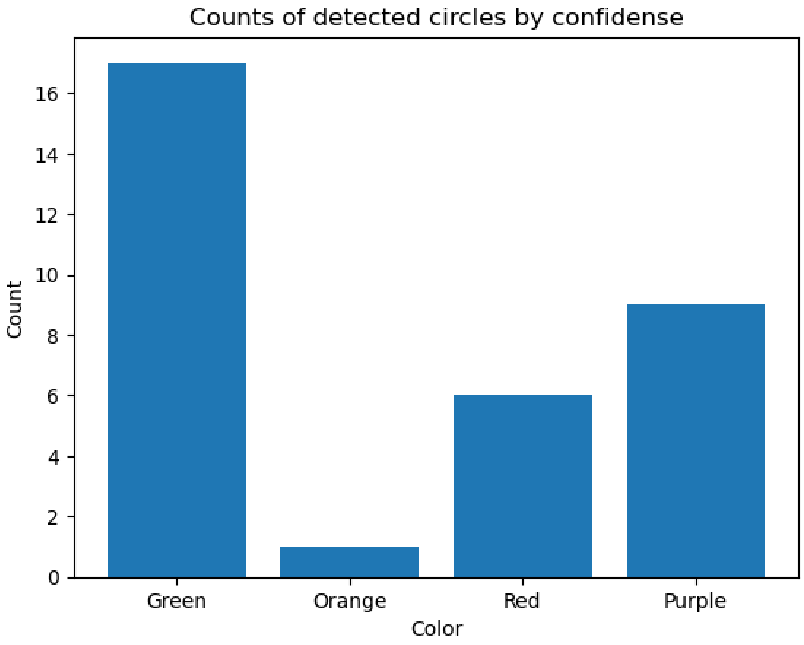


Fig. 4. Count of detected circles by confidence.

IV. Discussion

During the testing of the proposed algorithm, it became evident that it performs significantly better on images with a resolution larger than 1300x1300 pixels. However, it has limitations—it fails to work on images where the diameter of the oranges exceeds half of the screen size, or when the oranges are very small (less than 150 pixels). This limitation is attributed to the choice of the minimum separation value between circles in.

It was observed that these limitations can be effectively addressed by using higher image resolutions. In practical applications where image quality and consistent zoom levels are known and controlled, these shortcomings can be overcome. Additionally, analyzing luminance levels could be explored as a means to normalize images, particularly in cases where highly illuminated images pose challenges for orange detection.

The algorithm exhibits varying levels of performance due to inefficiencies within the circle detection process, making it computationally expensive and particularly evident in handling large and cluttered images. Further optimization is essential to enhance its efficiency and adaptability across different image scenarios.

V. Conclusion

The results derived from this algorithm demonstrate achieving an accuracy rate of approximately 72.4%. Result summarizes that purpose algorithm works well on lower fruit count. However, there remains a clear imperative for further enhancements. The algorithm's current limitations are evident in its inability to account for varying levels of ripeness in oranges. Moreover, the section responsible for orange detection lacks computational efficiency, thereby affecting program runtime.

Future research endeavors should prioritize the development of a more efficient algorithm for shape detection within images. Additionally, it is essential to optimize the color segmentation component to handle a broader range of brightness levels and mitigate issues related to clutter. These improvements will contribute to the advancement of automated fruit counting processes in agriculture and other related domains.

References

[1] Shankarpure, M. R.,& Patil, D. D. A Comprehensive Survey on Methods and Techniques for Automated Fruit Plucking. International Journal of Intelligent Systems and Applications in Engineering, 11(1), 156-168. 2023.

[2] Anuja Bhargava; Atul Bansal(2021). Fruits and vegetables quality evaluation using computer vision: A review, Journal of King Saud University - Computer and Information Sciences, Volume 33, Issue 3, Pages 243-257.

[3] Erdenee, B.; Ryutaro, T; Tana, G.(2010), Particular Agricultural Land Cover Classification Case Study Of Tsagaannuur, Mongolia. In: IEEE International Geoscience & Remote Sensing Symposium, 3194-3197.

[4] V.K. Tewari; A.K. Arudra; S.P. Kumar; V. Pandey; N.S. Chande (2013),Estimation of plant nitrogen content using digital image processing Int. Commission Agricu. Biosyst. Eng., pp. 78-86.

[5] M. Krishna; G. Jabert (2013) Pest control in agriculture plantation using image processing, IOSR J. Electron. Commun. Eng. (IOSR-JECE), pp. 68-74.

[6] J.K. Patil; R. Kumar (2011)Advances in image processing for detection of plant diseases, J. Adv. Bioinf. Appl. Res. ISSN, pp. 135-141.

[7] Hannan M.W.; Burks T.F.; Bulanon D.M.(2009), A Machine Vision Algorithm for Orange Fruit Detection, Agricultural Engineering International: the CIGREjournal, Vol-XI.

[8] Bulanon D.M.; Burks T.F.;Alchanatis V.(2009), Image Fusion of visible and thermal images for fruit detection, Biosystems Engineering, Vol-103, Issue-1, pages:12-22.

[9] Hannan M.W.; Burks T.F.; Bulanon D.M.(2009), A MachineVision Algorithm for Orange Fruit Detection, Agricultural Engineering International: the CIGRE journal, vol-XI,Pages:1-7.

[10] Leemans, V. and Destain, M.-F(2004), A real-time grading method of apple based on features extracted from defects, Journal of Food Engineering, vol.61, pp.83-89.

[11] Hayashi Shigehiko; Ota Tomohiko; Kubota Kotaro; Ganno Katsunobu and Kondo Naoshi (2005), Robotic Harvesting Technology for Fruit Vegetables in Protected Horticultural Production, Information and Technology for Sustainable Fruit and Vegetable Production FRUTIC , France.

[12] Bulanon D.M., Burks T.F. and Alchanatis V.(2008), Study of temporal variation in citrus canopy using thermal imaging for citrus fruit detection, Biosystems Engineering, Vol-101, Issue 2, Pages 161-171.

[13] Shasi Buluswar (2002), Models for Outdoor Color Vision, Doctoral dissertation, University of Massachusetts, Amherst.

[14] Bulanon D.M. ; Burks T.F.; Alcahnatis V.(2009) , Improving Fruit detection for robotic fruit harvesting", ISHS Acta Horticulturae 824: Internation Symosium on Application of Precision Agriculture for Fruits and Vegetables.

[15] Woo Chaw Seng and Seyed Hadi Mirisaee(2009), A New Method for Fruits Recognition System, MNCC Transactions on ICT, Vol. 1, No. 1.

[16] Blasco J.;Aleixos N.; Molto E.(2003), Machine Vision System for Automatic Quality Grading of Fruit, Biosystems Engineering , Vol-85, Issue 4, Pages-415-423.

[17] Fernández, L., Castillero, C. and Aguilera, J. M.(2005), An application of image analysis to dehydration of apple discs Journal of Food Engineering, vol.67, pp.185-193.

[18] Jimenez A.R., Ceres R., Pons J.L.(2000),A Survey of Computer Vision Methods for Locating Fruit on Trees, Transaction of the ASAE, Vol. 43(6), pages: 1911-1920.

[19] Borse, J.; Patil, D (2021). Tracking Keypoints from Consecutive Video Frames Using CNN Features for Space Applications. Tehnički glasnik, 15 (1), 11-17.

[20] Lokhande, Meghana P. and Dipti Durgesh Patil(2021),Trust Computation Model for IoT Devices Using Machine Learning Techniques, Proceeding of First Doctoral Symposium on Natural Computing Research. Lecture Notes in Networks and Systems, vol 169. Springer.

[21] Lokhande, Meghana P. and Dipti Durgesh Patil(2022),Object Identification in Remotely-Assisted Robotic Surgery Using Fuzzy Inference System, Demystifying Federated Learning for Blockchain and Industrial Internet.of Things, , pp. 58-73.

[22] A. Payne; K. Walsh; P. Subedi; and D. Jarvis(2013), Estimation of mango crop yield using image analysis – Segmentation method, Computers and Electronics in Agriculture, vol. 91, pp. 57–64.

[23] V. K. Yadav;S. Batham, A. K. Acharya, and R. Paul(2014), Approach to accurate circle detection: Circular Hough Transform and Local Maxima concept, in 2014 International Conference on Electronics and Communication Systems (ICECS), (Coimbatore), pp. 1–5, IEEE.

[24] Citrus Farming, https://www.usesfordiatomaceousearth.com/citrus-farming/ , 14 /10/2023.

[25] https://www.pexels.com/photo/orange-fruit-on-tree-3804878/, 23/10/2023.

[26] Stajnko, D., Lakota, M., Hocevar, M., 2004. Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging. Computers and Electronics in Agriculture 42, 31–34.

Towards a Framework for Elevating the Usage of eLearning Technologies in Higher Education Institutions

Naif Alzahrani[1], Hassan Alghamdi[2]
Department of Computer Engineering-School of Engineering, Al-Baha University[1]
Department of Computer Information Systems-School of CS and IT, Al-Baha University[2]
Al-Baha, Saudi Arabia[1, 2]

***Abstract*—Adopting eLearning technologies is no longer an option in Higher Education Institutions (HEIs) to support teaching and learning activities. However, despite steps taken by most of these institutions towards effective utilization of technologies, the current situation on the ground shows two significant challenges. First, the misalignment between institutions' strategies and the technical implementation of these technologies. Second, the scattered implementation and usage of eLearning technologies among stakeholders, which increases operational overhead due to the lack of a unified approach and usage procedures that promote optimal utilization of such technologies. This paper aims to introduce a framework for elevating the usage of eLearning technologies in HEIs. It guides the alignment between strategic goals and technology implementation for effective and progressive eLearning technology usage. Design science research methodology is adopted to guide the development of this framework. It drives the development process by first being aware of the problem from a real-life context and then proposing a solution. Principles from business and IT alignment and enterprise architecture are adopted to propose this framework, which is meant to be comprehensive to have eLearning technologies fit the institution's purpose while achieving strategic goals.**

Keywords*—eLearning; higher education; business and IT alignment; enterprise architecture

I. Introduction

Higher Education Institutions (HEIs) refer to universities, colleges, and other educational institutions that offer and deliver postsecondary education [1]. The teaching and learning activities carried out in these institutions are to prepare students for a specific profession [2]. Adopting Information and Communication Technologies (ICT) to support teaching and learning activities in these institutions is no longer an option. However, this requires continuous management and alignment of business needs with IT capabilities following HEIs' strategies. This strategic and operational alignment of business and IT has been a critical issue that faces organizations in different sectors, and the education sector is no exception.

This paper aims to develop a framework that assists in orchestrating alignment between the strategic vision and objective of HEIs with projects and activities related to using eLearning technologies. The paper is organized as follows: a literature review is carried out first to identify the gaps in IT investment and digital transformation in HEIs. Then, the business and IT alignment and the importance of employing enterprise architecture principles in HEIs are described, representing the theoretical and practical background to facilitate better management and alignment of business and IT resources in HEIs. The research methodology, which guides the research activities, is then explained. Then, the proposed framework developed according to design science research is introduced to contribute to the knowledge base. Finally, the conclusion and future work are explained.

II. Background

A. IT investments and Digital Transformation in HEIs

In the early days of adopting Information and Communication Technologies (ICT) in HEIs, they were mainly used for administrative and communication purposes rather than enhancing and supporting teaching and learning activities [3]. In recent years, however, ICT's investments and role in the education sector in general and in HEIs in particular have expanded enormously. These institutions increasingly invest in digital technologies to acquire sustainability, competitiveness, financial stability, innovation, stakeholders' satisfaction, teaching quality, institutional performance, and better ranking [4],[5]. New technologies are also adopted in these institutions to facilitate access to various online resources and distance education [6]. In a recent technical report, Morgan et al. [7] illustrate that advanced technologies in HEIs offer specialized web-conferencing tools, robotics process automation, access to artificial intelligence solutions, virtual tutoring platforms, advanced presentation technologies, and more solutions and focus on career support.

Embracing the education sector and institutions with advanced technologies is no longer a fancy option. HEIs, according to Pinho et al. [8] must adapt to the rapid changes in the technological environment as they have previously adapted to various social, political, and technical changes. According to the UNESCO National Commission [9], it is mandatory for HEIs, in particular, to proceed with technological enhancement to instruct and qualify people with the required knowledge that enables them to understand science and assists them in making the right decisions in terms of personal, professional and political choices.

The permeation of advanced technologies in every aspect of HEIs has forced these institutions to deal with holistic Digital Transformation (DT) planning, including all dimensions. A systematic literature review of DT in HEIs is carried out by Benavides et al. [10], and the findings reveal that DT in these institutions is still an emerging field and there is still a need for further research efforts to deal with the rapid changes in technology adoption in this sector. The authors state that although DT research papers related to HEIs have increased by 200% annually since 2016, the complex relationships between the actors involved in this technologically supported domain require further approaches to deal with a holistic transformation in HEIs, including business activities, processes, competencies, and models. Similarly, Durão et al. [11] pointed out that DT is not just about streamlining business processes and innovating new services. Still, it goes beyond to involve fundamental transformations in organizational procedures and capacity.

Realizing the benefits and Return on Investments (ROI) of ICT projects in HEIs can be overwhelming. Previous and recent studies have shown the challenges and failures associated with ICT investments in higher education institutions. For example, Kebritchi et al. [12] pointed out the challenges that hinder successful online course adoption in higher education. Issues are categorized in relation to learners, instructors, and the developed online content itself. Ejiaku [13] also pointed out that the absence of effective policies and leadership and the lack of IT professionals with the skills to analyze and manage IT projects have hindered the successful adoption of ICT in HEIs. Another study reveals that the low penetration of ICT systems among HEIs and the low level of IT literacy among students are considered significant challenges that affect the effective adoption of ICT in HEIs in a large population country like India [14]. Cloete [15] highlights that the complexity of IT itself and realizing its impacts on education varies from context to context, which can be the main challenges of embracing technology in education. Hatlevik [16] adds that Instructors' digital competence and beliefs about ICT are critical to effectively utilizing emerging technologies in teaching and learning institutions.

Although the business value of IT in many HEIs has been realized, it has not been effectively measured [17]. The Return on Investment (ROI) of IT capabilities in these institutions and this sector are largely not assessed [18]. Recently, policymakers and researchers have been demanding ROI calculation in higher education to see how students and HEIs perform [19]. This research does not aim to develop a tool that can assist in measuring the ROI or business value of IT in HEIs. However, it provides a clear representation of the misaligned components in these institutions and how they can be viewed and integrated through a framework to facilitate their value-adding to the business.

B. Business and IT Alignment in HEIs

Business and IT alignment is defined at strategic and operational levels according to the well-known model of alignment developed by Henderson et al. [20]. Reich et al. [21] define the alignment at the strategic level as "the degree to which the IT mission, objectives, and plans support and are supported by the business mission, objectives, and plans." On the other hand, Silvius et al. [22] define business and IT alignment at the operational level as "the degree to which IT applications, infrastructure and organization enable and support the business strategy and processes."

Most researchers and practitioners who studied and practiced the possibility of bridging the gap between business and IT agree that this harmony between these two domains positively impacts overall business performance [23]–[26]. Their work shows a consensus in their findings that organizations will perform well when IT resources, including hardware, software, IT skills, assets, and management, are aligned with business strategies and services. According to Liu et al. [27], it has become more critical in recent years to have the right IT with the right capabilities aligned with the right business requirements to have business value from IT investments.

The concept of alignment in academic and business sectors is one of the most frequently studied theories. Yet, many organizations are still misaligned and fail to take full advantage of IT [28], [29]. Business and IT alignment in HEIs represent a unique challenge due to the distinctive nature of these institutions. Decision-making in these institutions is shared, organizational culture among universities and colleges differs, and academic courses and research activities are independent [17]. According to Alghamdi et al. [30], business and IT alignment in the education sector has received the least attention in the business and IT alignment literature compared to other public and private sectors.

The isolation of IT planning when strategic business planning is carried out [31], the lack of top management awareness of the significant role of IT systems in organizations [32], the weak power sometimes of IT departments [31], and the rapid changing environment of business and IT [28], have all contributed to the ineffectiveness and misuse of IT in organizations. This has led many organizations in recent years to seek a holistic approach that provides an overarching view of the organization and its components in all its hierarchical layers (strategic, business, and technology), which can be found in the Enterprise Architecture theory and practice.

C. Enterprise Architecture in HEIs

Enterprise Architecture (EA) is a field of study that recently emerged to enhance the complexity and management of organizations and their business and IT domains to achieve strategic objectives and digital transformation [33], [34]. According to Gartner Group [35], "Enterprise Architecture is the process of translating business vision and strategy into effective enterprise change by creating, communicating, and improving key requirements, principles, and models that describe the enterprise's future state and enable its evolution.". In its simplest form, EA is described by Bernard [36] as the integration between the Strategy (S), Business (B), and Technology (T) layers of an enterprise, as illustrated in Fig. 1. The strategic layer at the top level is considered to be the main driver of an enterprise, while the business layer is the place where the source of requirements is elicited. The technology layer, the bottom layer, is where the provision of systems and technology to meet the business requirements to achieve strategic goals.

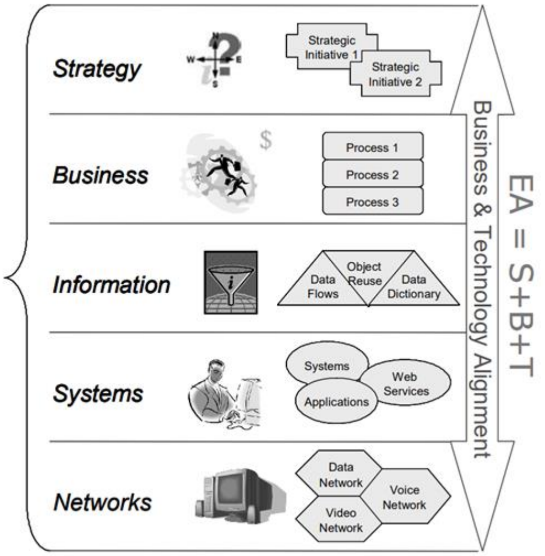


Fig. 1. The main layers of an enterprise from an enterprise architecture perspective [36].

During the last decade, there has been a growing interest in adopting EA principles and frameworks in HEIs. A recent systematic literature review carried out by Meutia et al. [37] shows that HEIs have adopted EA frameworks such as Zachman and TOGAF in the last decade for the following reasons: planning of IT infrastructure, integrating systems, developing appropriate strategic direction, achieving efficiency and effectiveness of available IT resources, ensuring standardized system development process, achieving competitive advantage through technology compared to other universities, reducing costs, supporting every business requirements, achieving better performance of systems, and finally achieving the mission and strategic goals related to the education process.

Another recent systematic mapping study of EA in the higher education domain is carried out by Bourmpoulias et al. [38]. Sixty articles were analyzed, and the authors pointed out that although the practice of EA in the education sector is widely accepted across the Americas, Asia, and Europe, the sector still faces major challenges in adopting EA principles and methodologies effectively. Among these challenges are the scarcity of EA practices in the education domain, the gap in the adoption and assessment of EA even in advanced education systems worldwide, and the focus on IT teams to lead the process and adoption of EA. The authors, on the other hand, highlight promising practices of EA in the education sector. Among these practices are adopting EA principles and frameworks in HEIs to assist in better business and IT planning, better business and IT alignment, effective change management, and better achievement of strategic goals and objectives. According to [10], EA practices and methodologies are enablers of DT in HEIs, where they can assist through systematic approaches the DT journey in these institutions.

III. Research Methodology

Design Science Research (DSR) is a research paradigm that guides a researcher in developing an artefact addressing a real-world problem and contributing to the knowledge base with the devised and evaluated solution [39]. DSR has received considerable attention in information systems research that focuses on the effective utilization of technologies in organizations [40], especially after the grounding work of [41]–[43].

Unlike the behavioral science paradigms that focus on developing and verifying theories to predict human or organizational behavior, the design science paradigm goes beyond human and organizational boundaries to develop new and innovative artefacts [42]. These artefacts are devised and evaluated following the DSR principles to solve a problem. According to Vaishnavi et al. [44], constructs, models, methods, instantiations, frameworks, architecture, design principles, and design theories are types of artefacts that can be the outcomes of a DSR project. A conceptual framework of DSR and its fundamental concepts is developed by Hevner et al. [42]. In their framework, three central components are emphasized: environment, IS research, and knowledge base. The environment specifies the boundaries of where the problem takes place involving people, organization, and technology and their underpinning components. The IS research and the knowledge base stand as a vehicle to address the issues identified in the environment based on the foundations of scientific theories and practices that support scientific IS research.

In reference [44], the authors describe a procedural workflow to carry out scientific research based on the developed DSR framework of Hevner et al. [42]. Fig. 2 shows the procedural flow of design science-based research that begins with awareness of the problem from a real-world context.

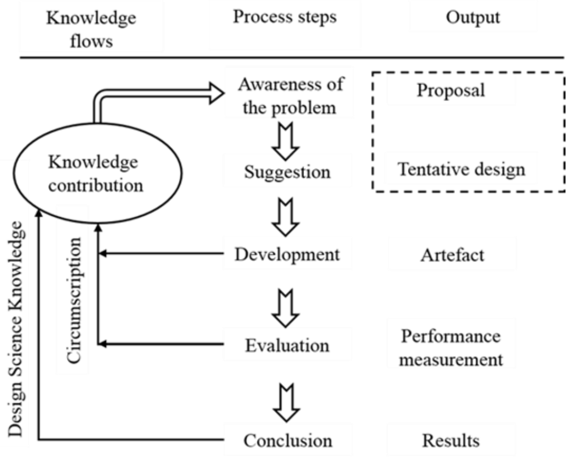


Fig. 2. A process model for DSR [44].

Then, an initial design is suggested to address the problem, which is driven by available relevant knowledge, either theoretical or practical. The initial design is then developed to derive the solution (i.e., the artefact). The resulting artefact is then evaluated through implementation to determine its ability to resolve identified problems. The conclusion then denotes the end of the procedural flow of the DSR, where the outcome itself is considered a contribution to the knowledge base. Some steps in the procedural flow can be iteratively carried out to enhance the outcomes of the DSR to provide a practical and effective solution to a real-world problem.

IV. Proposed Framework

Teaching and learning activities in their abstract form in a HEI consist of three main components: learning material (Courses), Instructors, and Students. eLearning technologies are constantly being procured and developed to improve these activities and maintain the essential components effectively involved in a positive relationship. This research aims to design a framework that supports HEIs to maintain an effective learning environment by elevating the usage of eLearning technologies. The alignment between operational activities in these institutions with strategic vision and objectives is considered along with principles and concepts from enterprise architecture theories while devising the framework. To achieve this, the DSR methodology is adopted as a scientific grounding to drive the development of the framework with an agile mindset that ensures continuous delivery.

A. *Awareness of the Problem*

Understanding the environment (people, organization, and technologies) is a primary activity in DSR to be aware of a real-world problem [42]. This assists in suggesting a solution that can address this problem and could be similar problems within the domain of study. The awareness of the problem in this research comes from threefold.

First, from the available literature on ICT investments and DT challenges in HEIs. Previous sections of this paper highlight related work that pointed out the challenges that face the effective utilization of advanced technologies in HEIs. The misalignment between eLearning technologies and HEIs' strategies can lead to a failure in achieving maximum utilization of these technologies to enhance overall teaching and learning activities. Such a challenge in the higher education sector has received the lowest attention in IT alignment research [30], [45]. A real example of misalignment practice that usually occurs in many HEIs is the isolation of implementing IT projects away from the actual needs of students and academic staff. In fact, the focus is shifted mainly to technological features rather than how these features can be utilized to improve teaching and learning quality. As a result, many HEIs have sets of scattered tools and systems with low utilization and an apparent absence of a unified ecosystem and effective utilization planning.

The literature also indicates that HEIs invest heavily in ICT to improve teaching and learning quality, one of their main common objectives. eLearning tools are employed in these institutions to achieve this quality in offering and delivering courses. However, utilizing eLearning technologies abstractly to deliver course materials has a low impact on the desired quality for three reasons. First, each instructor will develop their approach to using eLearning applications to achieve high quality in delivering course material and improving learning outcomes. As a result, the overall level of quality will vary between courses offered within one institution. Second, this quality cannot be measured due to the lack of measuring metrics. Therefore, it will be challenging to have an improvement plan since each instructor follows their way of employing eLearning tools. According to [46], quality practices in HEIs should be transformed for these institutions to remain effective. Third, learners will be confused and distracted between different tools used by other instructors. The utilization of eLearning technologies in HEIs without a unified framework that guides its adoption has a low chance of success due to the lack of alignment between institutions' vision and technological infrastructure implementation.

Second, the authors of this paper have worked for nine years as CIOs in a HEI in Saudi Arabia at (Al-Baha University). Therefore, they have carried out a requirement analysis process, which includes guided sessions to address the problem closely from the main stakeholders of the university (namely, students, academic staff, and top management), representing a real-life context. The outcomes assist in understanding the business and the actual needs of instructors and students to have effective learning activities and supporting eLearning tools during their academic journey. It also helps in being aware of the challenges, expectations, and beliefs of decision-makers towards making investments and supporting the adoption of eLearning technologies. Being CIOs at the university for several years, the authors are entirely aware of the technical capabilities of the university and its misalignment with its strategic scene.

Third, commonly utilized technologies and techniques in HEIs in Saudi Arabia and worldwide are investigated, including LMS, IT infrastructure, online course development process, and the eLearning platform's operation models. This allows for a better understanding of the environment and available ICT solutions in higher education.

Being aware of the environment and these challenges in HEIs, the proposed framework is suggested to provide a practical guideline to systematically overcome these challenges to elevate the usage of eLearning technologies in HEIs to effectively realize strategic goals through guidelines from enterprise architecture and alignment theory.

B. *Developed Framework*

Environmental analysis has guided the development of a set of artefacts and a framework to address the defined problem. The first artefact is shown in Fig. 3, which is a model that includes the four main key interactive components in an HEI domain that adopts eLearning technologies. These components are the Learner, Instructor, Course, and Platform, which supports teaching and learning activities (LICP). They must be aligned to maintain an effective relationship between them that result in generating business value of IT investment.

The model in Fig. 3 also illustrates the major interactive activities that are directly supported by a technological platform that assists teaching and learning activities. The learning platform should always afford the required functionalities from learners and instructors, following best practices and quality standards to ensure a successful learning journey. Therefore, the effective utilization of eLearning technologies should positively impact other activities. For instance, providing a unified process for the course development lifecycle using eLearning tools implemented on a centralized platform will support instructors in developing and enhancing their courses to meet defined quality requirements. Moreover, such unified processes and platforms will support the creation of course libraries that can be shared with all instructors and provide students with a high-quality eLearning experience.

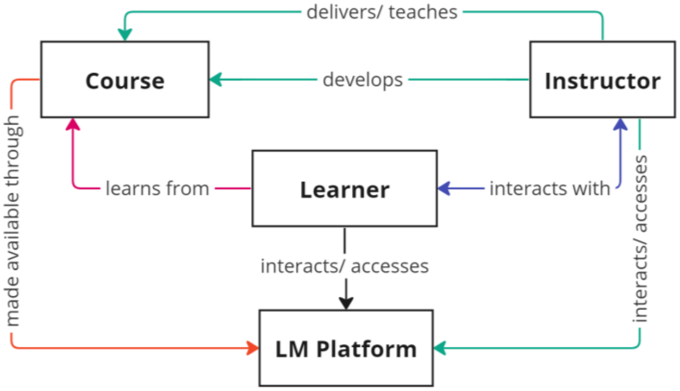


Fig. 3. Main interactive components in a HEI (LICP).

The second artefact developed based on the environmental analysis is shown in Fig. 4. It is another model that considers the principles of EA to draw the relationship between LICP components identified in the first model.

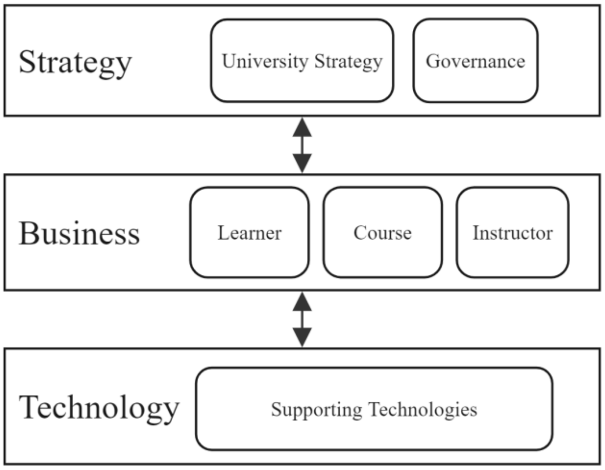


Fig. 4. LICP components from an EA point of view.

The model in Fig. 4 highlights, from an EA perspective, the need to have a comprehensive overview of the HEI that integrates supporting technologies with main business components to realize an overall institution strategy. For instance, integrating Learning Management Systems (LMS), eTest, virtual labs, and virtual classes and providing a usage guideline would support business actors (instructors and students) to effectively communicate during the teaching and learning journey, having an impact on the quality of education that the institution seeks as a strategic objective.

The model also highlights the need to define an ecosystem combining components to drive the institution's implementation, adoption, or elevation of technologies. This ecosystem is not necessarily technology-centric but rather a customized set of components driven from assessing an institution’s capabilities to assist institutions in moving from the current view (as-is) to the targeted future view (to-be). An eLearning ecosystem, for example, includes policies, procedures, tools, actors, and infrastructure components such as Learning Management System (LMS), Student Information System (SIS) integration, and students email integration, labs, and eTest centers that should be harnessed according to institution’s capability to elevate the usage of supporting technologies.

Based on these artefacts, the proposed framework is developed, which consists of seven iterative stages as illustrated in Fig. 5. It aims to assist HEIs in elevating the usage of eLearning technologies by adopting them in a systematic way that ensures the involvement of related organizational components from alignment and EA perspectives. This way, improving the teaching and learning journey for the main stakeholders in this sector (students and instructors) through a unified structure is prioritized while keeping LICP components aligned with the institution’s strategic vision and goals through governance and an agile mindset supported by EA principles.

Each stage defined in the framework is considered a milestone in utilizing or elevating the usage of eLearning technologies in HEIs through a project roadmap. Each iteration will focus on a specific goal to promote progressive quality improvement in HEIs. Table I illustrates a summary of activities that should be carried out in each stage, along with expected outcomes.

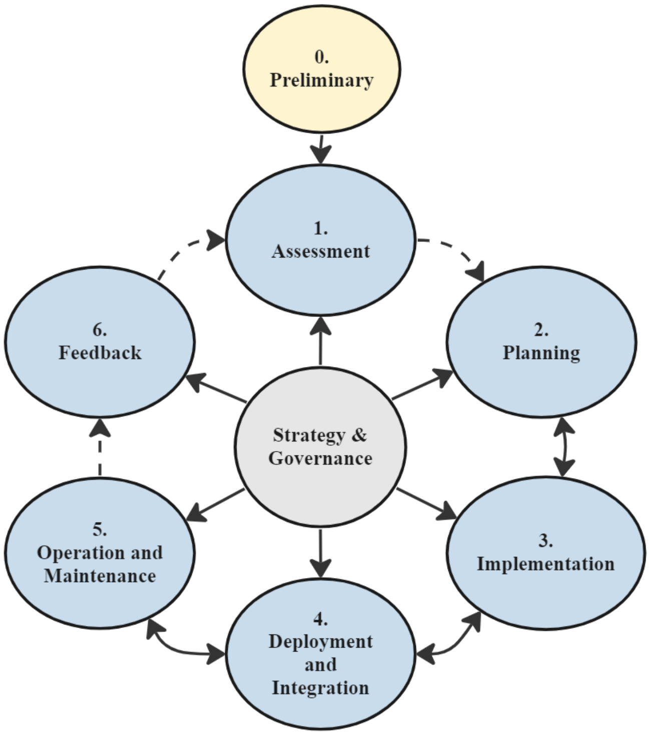


Fig. 5. Proposed framework for elevating the usage of eLearning technologies in HEIs.

TABLE I. STAGES OF THE PROPOSED FRAMEWORK

Stage	Main Activities	Main Outcomes
0 - Preliminary	● Form a core team. ● Identify eLearning goals. ● Select project management tools.	● Formed team members with specified roles and responsibilities. ● Identified eLearning objectives. ● Set up project backlog tool.
1 - Assessment	● Gap identification. ● Competencies assessments for LICP components.	● SWOT analysis outcomes. ● Requirements specification to reduce the gap between (As-IS) and (To-Be) state. ● List of goals aligned with the institution's strategy.
2 - Planning	● Design eLearning projects. ● Specify project dependencies. ● Identify iterations (sequential and parallel).	● List of outcomes that must be delivered for each project. ● Projects roadmap. ● Projects implementation plan.
3 - Implementation	● Incremental implementation. ● Update project backlog	● Project progress reports. ● Tested and verified project ready to be integrated into an eLearning ecosystem.
4 - Deployment and Integration	● Update the eLearning ecosystem. ● Integration with existing components.	● Updated ecosystem according to project scope.
5 - Operation and maintenance	● Maintain ecosystem operation. ● Collect predefined metrics, e.g., course material usage	● Process for proactive action to avoid any technical issues. ● Updated ecosystem operation manual.
6 - Feedback	● Collect feedback. ● Feedback analysis.	● Feedback results. ● Updated metrics.

Institution's strategy and governance are at the heart of this framework. The former guides the following assessment and planning stages, and the outcomes of all stages also impact it, while the latter controls the procedures and activities defined in each stage to ensure compliance with the requirements and regulations. Effective governance should ensure the missing alignment between the institution's layers (strategy, business, and technology). Governance also controls change management activities that support promoting eLearning culture and adopting new modes of teaching and learning. Change management also provides a systematic approach to implementing new projects approved by top management and helps to guide the activities that will be performed in the following stages of the framework.

The preliminary stage is the starting point of the proposed framework, which focuses on two activities. First, a core team will be formed to implement the following stages. Second, identify eLearning-related objectives based on HEI's strategies, which will be prepared to go through the iterative process of implementing subsequent framework stages. In this stage, the team will adopt an agile project management mindset in planning and executing each iteration. As a result, a backlog of eLearning projects and their objectives will be the outcomes of this stage that need to be approved by top management in terms of prioritization to ensure its alignment with other initiatives and projects at the HEI.

In the assessment stage, the core team uses a set of tools based on the outcomes of the preliminary stage to identify gaps between the current state and the targeted state of the eLearning tools adoption perspective, such as SWOT analysis (Strengths, Weaknesses, Opportunities, and Threats) and academic staff competencies' evaluation. etc., This stage triggers the alignment process between the institution's strategies and available eLearning capabilities. The outcomes of this stage should be a list of requirements that, once attained, should transform the institution from state (as-is) to state (to-be), realizing the institution's strategic objectives according to available and obtained capabilities.

Defining an eLearning ecosystem consisting of a set of components supports the improvement of eLearning capabilities at the institution, and this ecosystem works as an enabler for projects' implementations identified in stage two. Each project related to eLearning will be designed to use and improve this ecosystem. Integrating all components in this ecosystem is intended to create a single and comprehensive learning environment that combines all LICP dimensions for better and effective interactions. This also helps collect feedback from LICP for continuous improvement according to the institution's strategic objectives.

The planning stage uses the outcomes of previous stages to plan eLearning projects based on the targeted state of the institution. An agile project management mindset should be employed in this stage to ensure continuous delivery of the eLearning functionality specified in each project. As mentioned, the framework is iterative and progressive and can be achieved by running multiple projects simultaneously. The core team should study projects' dependencies and design an implementation roadmap that keeps all projects aligned. Such an approach guarantees fast delivery and continuous improvement since the stages from three to six can be executed independently for each project.

The operational core team can coordinate and achieve activities in stages three, four, and five. In this, multiple projects can be running concurrently through these stages. For example, one project may focus on online course development based on a Course Development Lifecycle (CDLC) methodology, while another project may implement new functionality in the eTest center. These two projects represent valuable components of an eLearning ecosystem that aims to achieve an institution's strategic goals through a systematic approach. The role of project managers and project management tools is critical to successfully implementing defined projects since they help identify dependencies and implement the roadmap.

The feedback stage is crucial because all eLearning projects must be aligned with the institution's vision and objectives and meet the needs and expectations of students and instructors, who are the primary stakeholders of these projects. Through this framework, feedback is constantly collected and used as input for the next iteration of project implementation. This guarantees effective engagement and utilization of all features developed and applied to the eLearning environment.

V. Conclusion and Future work

Higher education institutions invest heavily in supporting technologies to achieve better outcomes regarding the quality of delivered education. However, the alignment of these technologies with higher education institutions' strategies and the exact needs of students and instructors in this sector is questionable. The absence of measuring tools to assess this alignment and the continuous rapid investments in advanced technologies in higher education require efforts to evaluate the business value of IT in this sector.

A framework for elevating the usage and adoption of eLearning technologies in the higher education sector has been proposed in this paper. Design science research is adopted as a research methodology to guide its activities, which is found suitable for the problem-based nature of this research. Being aware of the problems faced in this sector regarding effective IT utilization and adopting principles and practices from enterprise architecture and business and IT alignment, a set of artefacts and a framework are developed. These principles and practices support having a holistic view of institutions that guide the adoption of advanced technologies through systematic approaches that ensure effective business and IT alignment for better utilization.

The implementation of the framework has several advantages. First, it enforces a standardized approach to utilizing and leveraging the usage of eLearning technologies that directly impact the quality of teaching. Second, it offers a baseline for basic metrics for measuring the institution's quality improvement progress in educational technologies. Third, due to the unified implementation approach and progressive improvement, instructors' competencies should be improved, and students' academic achievement should be enhanced accordingly. Moreover, the iterative stages of the proposed framework support the inevitable adoption and adaptation of any advanced eLearning technologies.

The proposed framework is applied at AL-Baha University to evaluate its viability in elevating the usage of eLearning technologies and establishing a unified learning environment governed by a set of processes. The application and findings are published in [47].

Acknowledgments

The outcomes of this research are part of the authors' cooperation in implementing a digital transformation plan at the Deanship of eLearning and Information Technology (DeIT) at Al-Baha University. The authors thank the deanship of scientific research at Al-Baha University, Kingdom of Saudi Arabia, for sponsoring this project in 1442H with a reference number (1442/2). Sincerest and deep appreciation go to the eLearning team at Al-Baha University.

References

[1] G. Gerón-Piñón, P. Solana-González, D. Pérez-González, and S. Trigueros-Preciado, "Information System Projects for Higher Education Management: Challenges for Latin American Universities," in Higher Education and the Evolution of Management, Applied Sciences, and Engineering Curricula, IGI Global, 2019, pp. 120–150.

[2] M. Klumpp and U. Teichler, "German Fachhochschulen: Towards the end of a success story?," in Non-university higher education in Europe, Springer, 2008, pp. 99–122.

[3] R. B. Kvavik, "Convenience, communications, and control: How students use technology," Educating the net generation, vol. 1, no. 2005, pp. 1–7, 2005.

[4] Y.-S. Tsai et al., "Learning analytics in European higher education—Trends and barriers," Comput Educ, vol. 155, p. 103933, 2020.

[5] A. F. Teixeira, M. J. A. Gonçalves, and M. de L. M. Taylor, "How higher education institutions are driving to digital transformation: A case study," Educ Sci (Basel), vol. 11, no. 10, p. 636, 2021.

[6] K. V. Z. Caliari, M. A. Zilber, and G. Perez, "Tecnologias da informação e comunicação como inovação no ensino superior presencial: uma análise das variáveis que influenciam na sua adoção," REGE-Revista de Gestão, vol. 24, no. 3, pp. 247–255, 2017.

[7] G. Morgan et al., "Top technology trends in higher education for 2022." Technical report, Gartner, 2022.

[8] C. Pinho, M. Franco, and L. Mendes, "Exploring the conditions of success in e-libraries in the higher education context through the lens of the social learning theory," Information & Management, vol. 57, no. 4, p. 103208, 2020.

[9] "Investing in science, technology and research - Science for a sustainable future - Themes - UNESCO National Commission," UNESCO. Accessed: Oct. 07, 2023. [Online]. Available: https://unescoportugal.mne.gov.pt/pt/temas/ciencia-para-um-futuro-sustentavel/investir-na-ciencia-tecnologia-e-investigacao

[10] L. M. C. Benavides, J. A. Tamayo Arias, M. D. Arango Serna, J. W. Branch Bedoya, and D. Burgos, "Digital transformation in higher education institutions: A systematic literature review," Sensors, vol. 20, no. 11, p. 3291, 2020.

[11] N. Durão, M. J. Ferreira, C. S. Pereira, and F. Moreira, "Current and future state of Portuguese organizations towards digital transformation," Procedia Comput Sci, vol. 164, pp. 25–32, 2019.

[12] M. Kebritchi, A. Lipschuetz, and L. Santiague, "Issues and challenges for teaching successful online courses in higher education: A literature

review," Journal of Educational Technology Systems, vol. 46, no. 1, pp. 4–29, 2017.

[13] S. A. Ejiaku, "Technology adoption: Issues and challenges in information technology adoption in emerging economies," Journal of International Technology and Information Management, vol. 23, no. 2, p. 5, 2014.

[14] U. K. Pegu, "Information and communication technology in higher education in india: Challenges and opportunities," International Journal of Information and Computation Technology, vol. 4, no. 5, pp. 513–518, 2014.

[15] A. L. Cloete, "Technology and education: Challenges and opportunities," HTS: Theological Studies, vol. 73, no. 3, pp. 1–7, 2017.

[16] O. E. Hatlevik, "Examining the relationship between teachers' self-efficacy, their digital competence, strategies to evaluate information, and use of ICT at school," Scandinavian Journal of Educational Research, vol. 61, no. 5, pp. 555–567, 2017.

[17] J. A. Pirani and G. Salaway, "Information technology alignment in higher education," Educause Center for Applied Research, pp. 1–10, 2004.

[18] R. Abel, "Innovation, adoption, and learning impact: Creating the future of IT," EducAusE review, vol. 42, no. 2, pp. 13–14, 2007.

[19] K. Blagg and E. Blom, "Evaluating the Return on Investment in Higher Education: An Assessment of Individual-and State-Level Returns.," Urban Institute, 2018.

[20] J. Henderson and N. Venkatraman, Strategic alignment: a model for organizational transformation via information technology. Oxford University Press New York, 1990.

[21] B. H. Reich and I. Benbasat, "Measuring the linkage between business and information technology objectives," MIS quarterly, pp. 55–81, 1996.

[22] A. J. Silvius, B. De Waal, and J. Smit, "Business and IT alignment; answers and remaining questions," Pacis 2009 Proceedings, p. 44, 2009.

[23] T. Coltman, P. Tallon, R. Sharma, and M. Queiroz, "Strategic IT alignment: twenty-five years on," Journal of Information Technology, vol. 30. Springer, pp. 91–100, 2015.

[24] J. E. Gerow, J. B. Thatcher, and V. Grover, "Six types of IT-business strategic alignment: an investigation of the constructs and their measurement," European Journal of Information Systems, vol. 24, no. 5, pp. 465–491, 2015.

[25] H.-T. Wagner and T. Weitzel, "Operational IT business alignment as the missing link from IT strategy to firm success," AMCIS 2006 Proceedings, p. 74, 2006.

[26] A. A. Yayla and Q. Hu, "The impact of IT-business strategic alignment on firm performance in a developing country setting: exploring moderating roles of environmental uncertainty and strategic orientation," European Journal of Information Systems, vol. 21, no. 4, pp. 373–387, 2012.

[27] K. Liu and W. Li, Organisational semiotics for business informatics. Routledge, 2014.

[28] A. Ullah and R. Lai, "A systematic review of business and information technology alignment," ACM Transactions on Management Information Systems (TMIS), vol. 4, no. 1, pp. 1–30, 2013.

[29] H. Aggarwal, "Contemporary Research Issues in Business–IT Alignment," Digitising Enterprise in an Information Age, pp. 3–15, 2021.

[30] H. Alghamdi and L. Sun, "Business and IT alignment in higher education sector," International Journal of Technology and Engineering Studies, vol. 3, no. 1, pp. 1–8, 2017.

[31] M. Tarafdar and T. S. Ragu-Nathan, "Business–Information Systems Alignment: Taking Stock and Looking Ahead," in Planning for Information Systems, Routledge, 2015, pp. 46–79.

[32] J. W. Weiss and D. Anderson, "Aligning technology and business strategy: Issues & frameworks, a field study of 15 companies," in 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, IEEE, 2004, pp. 10-pp.

[33] J. Lapalme, A. Gerber, A. Van der Merwe, J. Zachman, M. De Vries, and K. Hinkelmann, "Exploring the future of enterprise architecture: A Zachman perspective," Comput Ind, vol. 79, pp. 103–113, 2016.

[34] R. Van de Wetering, S. Kurnia, and S. Kotusev, "The role of enterprise architecture for digital transformations," Sustainability, vol. 13, no. 4. MDPI, p. 2237, 2021.

[35] Gartner Group, "Enterprise Architecture," https://www.gartner.com/en/information-technology/glossary/enterprise-architecture-ea.

[36] S. Bernard, "Using enterprise architecture to integrate strategic, business, and technology planning," Journal of Enterprise Architecture, vol. 2, no. 4, pp. 11–28, 2006.

[37] N. S. Meutia, E. Sulistiyani, R. P. N. Budiarti, and R. Sari, "Enterprise Architecture Framework in Higher Education: Systematic Literature Review," Applied Technology and Computing Science Journal, vol. 5, no. 2, pp. 112–118, 2022.

[38] S. Bourmpoulias and K. Tarabanis, "A systematic mapping study on Enterprise Architecture for the Education domain: Approaches and Challenges," in 2020 IEEE 22nd Conference on Business Informatics (CBI), IEEE, 2020, pp. 30–39.

[39] A. Hevner, S. Chatterjee, A. Hevner, and S. Chatterjee, "Introduction to design science research," Design research in information systems: theory and practice, pp. 1–8, 2010.

[40] C. Fischer, R. Winter, and F. Wortmann, "Design theory," Business & Information Systems Engineering, vol. 2, pp. 387–390, 2010.

[41] S. T. March and G. F. Smith, "Design and natural science research on information technology," Decis Support Syst, vol. 15, no. 4, pp. 251–266, 1995.

[42] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," Management Information Systems Quarterly, vol. 28, no. 1, p. 6, 2008.

[43] J. G. Walls, G. R. Widmeyer, and O. A. El Sawy, "Building an information system design theory for vigilant EIS," Information systems research, vol. 3, no. 1, pp. 36–59, 1992.

[44] V. K. Vaishnavi and W. Kuechler, Design science research methods and patterns: innovating information and communication technology. Crc Press, 2015.

[45] N. Alshareef, S. M. Elakeil, and A. M. Maatuk, "A Framework of Information Systems Reference Model for Higher Education Institutions," in 2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), IEEE, 2023, pp. 392–396.

[46] B. Alzahrani, H. Bahaitham, M. Andejany, and A. Elshennawy, "How ready is higher education for quality 4.0 transformation according to the LNS research framework?," Sustainability, vol. 13, no. 9, p. 5169, 2021.

[47] H. Alghamdi and N. Alzahrani, "Evolving Adoption of e-Learning Tools and Developing Online Courses: A Practical Case Study from Al-Baha University, Saudi Arabia," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. (in press), 2023.

A Novel Paradigm for IoT Security: ResNet-GRU Model Revolutionizes Botnet Attack Detection

Jyotsna A[1], Mary Anita E.A[2]

Dept. of Computer Science and Engineering, Christ (Deemed to be University), Bangalore, India[1]
Dept. of Computer Science and Engineering, RSET, Kochi, India[1]
Dept. of Computer Science and Engineering, Christ (Deemed to be University), Bangalore, India[2]

***Abstract*—The rapid proliferation of the Internet of Things (IoT) has engendered substantial security apprehensions, chiefly due to the emergence of botnet attacks. This research study delves into the realm of Intrusion Detection Systems (IDS) by leveraging the IoT23 dataset, with a specific emphasis on the intricate domain of IoT at the network's edge. The evolution of edge computing underscores the exigency for tailored security solutions. An array of statistical methodologies, encompassing ANOVA, Kruskal-Wallis, and Friedman tests, is systematically employed to illuminate the evolving trends across multiple facets of the study. Given the intricacies entailed in feature selection within edge environments, Chi-square analyses, Recursive Feature Elimination (RFE), and Lasso-based techniques are strategically harnessed to unearth meaningful feature subsets. A meticulous evaluation encompassing 19 classifiers, meticulously selected from both machine learning (ML) and deep learning (DL) paradigms, is rigorously conducted. Initial findings underscore the potential of the Gated Recurrent Unit (GRU) model, especially when coupled with intrinsic lasso-based feature selection. This promising outcome catalyzes the formulation of an ensemble approach that harnesses multiple LassoCV models, aimed at amplifying feature selection proficiency. Furthermore, an optimized ResNet-GRU model emerges from the fusion of the GRU and ResNet architectures, with the objective of augmenting classification performance. In response to mounting concerns regarding data privacy at the edge, a resilient federated learning ecosystem is meticulously crafted. The seamless integration of the optimized ResNet-GRU model into this framework facilitates the employment of FedAvg, a widely acclaimed federated learning methodology, to adeptly navigate the intricacies associated with data sharing challenges. A comprehensive performance evaluation is undertaken, wherein the ResNet-GRU model is benchmarked against FedAvg and a diverse array of other federated learning algorithms, including FedProx and Fed+. This extensive comparative analysis encompasses a spectrum of performance metrics and processing time benchmarks, shedding comprehensive light on the capabilities of the model.**

Keywords*—Internet of things; federated learning; Gated Recurrent Neural Networks; Long Short Term Memory (LSTM)

I. INTRODUCTION

The Internet of Things (IoT) has transformed device connectivity, bringing benefits and challenges in anomaly detection. IoT anomalies can stem from various factors like environmental changes, cybersecurity breaches, or device failures [1]. Detecting and understanding these anomalies are vital for ensuring dependability, security, and performance. Anomalies can disrupt operations, compromise data security, or invade privacy, necessitating proactive identification. Specialized methods, including artificial intelligence, machine learning, and statistical analysis, are essential for anomaly detection. IoT devices are susceptible to cybersecurity threats like unauthorized access and data breaches, involving atypical network traffic, unusual access patterns, or suspicious user behavior. Detecting these anomalies is crucial for preventing security breaches. Additionally, individual IoT devices may exhibit unexpected behavior due to software, firmware, or hardware issues [2]. Promptly identifying and resolving these device anomalies is essential for maintaining device reliability.

Addressing IoT anomalies involves various techniques, including artificial intelligence, machine learning, statistical analysis, and anomaly detection algorithms. These methods aim to establish normal behavior, detect deviations, and trigger appropriate responses. Machine learning and deep learning are particularly effective due to their ability to analyze vast datasets and identify patterns, enhancing IoT system safety and reliability [3]. In this article, we explore how machine learning and deep learning are applied in IoT anomaly detection. Supervised learning is used with labeled datasets, training models on historical data to recognize normal behavior patterns and identify anomalies in real-time data. Algorithms like decision trees, support vector machines (SVM), random forests, and gradient boosting are employed [4]. In cases of limited or unlabeled data, unsupervised learning becomes essential. It identifies anomalies by analyzing data structures and trends, utilizing techniques such as clustering to group data and detect anomalies as outliers [5].

Autoencoders are a type of neural network, excel in IoT anomaly detection by reducing input data dimensionality and detecting anomalies through reconstruction errors. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are ideal for handling sequential IoT data. Generative Adversarial Networks (GANs) are suitable for anomaly detection, as they can simulate complex IoT data distributions [6]. Federated learning is a decentralized approach for IoT anomaly detection that preserves data privacy [7]. IoT devices locally train models, transmitting only model updates, reducing the need for data transfer to a centralized server. This approach is beneficial for scenarios with limited bandwidth or intermittent connectivity, enhances model stability, and accommodates device-specific constraints [8]. However, it introduces communication overhead and security concerns [9].

The proposed work analyzes the IoT23 dataset, focusing on botnet attacks, using statistical tests and three feature selection approaches (filter, wrapper, embedded). Fifteen classifiers, including machine learning and deep learning models, are evaluated, with the best-performing one being the GRU model with embedded lasso-based feature selection. An ensemble of LassoCV models and ResNet architecture further improves feature selection and classifier performance. To address privacy concerns, a federated learning environment is established, and the optimized ResNet-GRU model is deployed and compared with existing federated learning algorithms, considering various metrics and processing time.

- The study focuses on Intrusion Detection Systems (IDS) within the context of the Internet of Things (IoT) at the network's edge, addressing heightened security concerns due to botnet attacks.
- Edge computing's evolution necessitates customized security solutions, and this research endeavors to provide them.
- Various statistical methodologies, including ANOVA, Kruskal-Wallis, and Friedman tests, are employed to reveal evolving trends in IoT security at the network's edge.
- Innovative feature selection techniques such as Chi-square analyses, Recursive Feature Elimination (RFE), and Lasso-based methods are applied to navigate the complexities of feature selection in edge environments.
- The study rigorously evaluates 19 classifiers from both machine learning (ML) and deep learning (DL) domains, with a particular focus on the Gated Recurrent Unit (GRU) model, which shows promise in conjunction with lasso-based feature selection.
- A novel ensemble approach harnessing multiple LassoCV models is developed to enhance feature selection efficiency.
- The introduction of an optimized ResNet-GRU model, combining GRU and ResNet architectures, aims to improve classification performance.
- To address data privacy concerns at the edge, a resilient federated learning ecosystem is created, integrating the optimized ResNet-GRU model and employing FedAvg, a widely acclaimed federated learning methodology.
- Comprehensive performance evaluation includes benchmarking against FedAvg and various other federated learning algorithms, such as FedProx and Fed+, covering a wide range of performance metrics and processing time benchmarks.

II. Related Works

This section presents the related works carried out by several research scholars in the area of anomaly detection using deep learning and machine learning with the aid of federated learning.

Brett Weinger et al. [10] employed Federated Learning (FL) for collaborative mobile and IoT projects but faced technological challenges. Distributing ML training across devices reduced prediction accuracy compared to centralized learning. Limited data access led to issues like constrained local ML models and class imbalances due to diverse event contributions. They addressed these challenges with data augmentation, resulting in a significant 22.9% performance improvement in IoT anomaly detection across three datasets.

Zhuotao Lian et al. [11] enhanced IoT anomaly detection while addressing security concerns. They proposed a distributed federated learning approach using neural networks, as traditional methods proved inaccurate. This technique improved detection accuracy while safeguarding locally stored data through decentralized learning, eliminating central failure points and raw data flow. Simulations using the IoT23 dataset validated its effectiveness, showcasing the promise of distributed learning for secure and accurate IoT anomaly detection, nearly matching centralized federated learning in performance.

Truong Thu Huong et al. [12] developed the FedeX architecture for efficient distributed anomaly detection in IoT-based Industrial Control Systems (ICSs) for Smart Manufacturing. FedeX outperformed 14 other methods on various detection measures, offering rapid learning, lightweight deployment, and interpretability. It allows real-time edge deployment with 7.5 minutes of training and 14% memory use, enhancing Smart Manufacturing practices. Explainable AI (XAI) to improve model interpretability, helping experts make confident decisions.

Subir Halder et al. [13] developed Hawk, an anomaly detection system for LoRa-enabled IIoT networks to address cybersecurity challenges. Hawk uses unique Carrier Frequency Offset (CFO) measurements to create device "fingerprints" and detect suspicious behavior. Employing federated learning, Hawk outperformed other systems by over 8% in detection accuracy and demonstrated high resilience against cyberattacks, reducing storage overhead by 40%. It's an effective solution for securing LoRa-enabled IIoT networks against novel threats.

Xabier Sáez-de-Cámara et al. [14] addressed IoT cybersecurity challenges in their study. They proposed a system using unsupervised models for network intrusion detection in large and diverse IoT and IIoT deployments. To overcome issues like network overhead and heterogeneity, they leveraged Federated Learning (FL) for cooperative training. Their architecture, tested on a simulated network with 100 nodes and subjected to real-world attacks, demonstrated efficient and robust intrusion detection for large-scale IoT and IIoT environments.

Huong Thu Truong et al. [15] developed a scalable anomaly detection system for continuously operating Industrial Control Systems (ICS) in smart manufacturing and IIoT. Their system combines Federated Learning, Autoencoder, and Transformer, with a Fourier mixing sublayer for improved performance. It offers rapid training within minutes, is lightweight with low computational and memory requirements, and minimizes communication costs. Compared to existing methods, it reduces training time by 50% to 1200 seconds, adapting to changing conditions and mitigating false positives in ICS data patterns, ensuring robust anomaly detection for smart manufacturing.

Jiamin Fan et al. [16] developed Score-VAE, a novel root cause analysis method for IoT anomaly detection systems. It addresses the challenge of distinguishing false positives from malicious attacks. Score-VAE combines the training and testing schemes of the VAE network within the federated learning (FL) architecture, resulting in improved generalization, learning, collaboration, and privacy protection. It effectively identifies the sources of anomaly detection alarms in real-world IoT data, outperforming standard approaches and enhancing the accuracy of root cause analysis in IoT anomaly detection.

Ali Raza et al. [17] introduced AnoFed, a novel federated framework for anomaly detection in digital healthcare, particularly in ECG analysis. To overcome limitations in threshold selection and privacy concerns in centralized machine learning, they combined transformer-based AE and VAE with Support Vector Data Description (SVDD). AnoFed enhances privacy, improves interpretability, and facilitates adaptive anomaly detection. Experiments in ECG anomaly detection demonstrated its effective performance with low computational costs. AnoFed's efficiency and privacy-preserving capabilities make it a valuable solution for digital healthcare applications, suitable for deployment on low-powered edge devices.

J. Jithish et al. [18] conducted a technical study in the past, focusing on anomaly detection in the smart grid using Federated Learning (FL). Anomaly detection is crucial for identifying energy theft, cyberattacks, and excessive power usage. In this approach, smart metres locally train machine learning models without sending data to a centralised server. Smart metres download a global model from the server for on-device training, and after local training, upload model parameters to fine-tune the global model, protected by the SSL/TLS protocol. Experiments on industry-standard datasets demonstrated that FL models matched the accuracy of centralised ML models while preserving individual privacy. The research showcased the efficiency of FL-based models in terms of memory, CPU utilization, bandwidth, and power consumption at edge devices, making them suitable for deployment in resource-constrained settings like smart metres in the smart grid.

III. MATERIALS AND METHODS

This study focuses on IoT network botnet attack detection using feature selection and classification techniques. It employs three feature selection methods (filter, wrapper, and embedded) to reduce dataset dimensionality. Out of 15 classifiers tested, the GRU model with embedded lasso-based feature selection emerges as the top performer [19]. To enhance detection capabilities, an ensemble approach is applied, incorporating 10 LassoCV models. Additionally, optimization with the ResNet architecture is employed to improve detection accuracy and convergence speed by addressing the vanishing gradient problem. This comprehensive approach aims to provide more effective and efficient botnet attack detection in IoT networks using the IoT23 dataset. The detailed architecture is presented in Fig. 1.

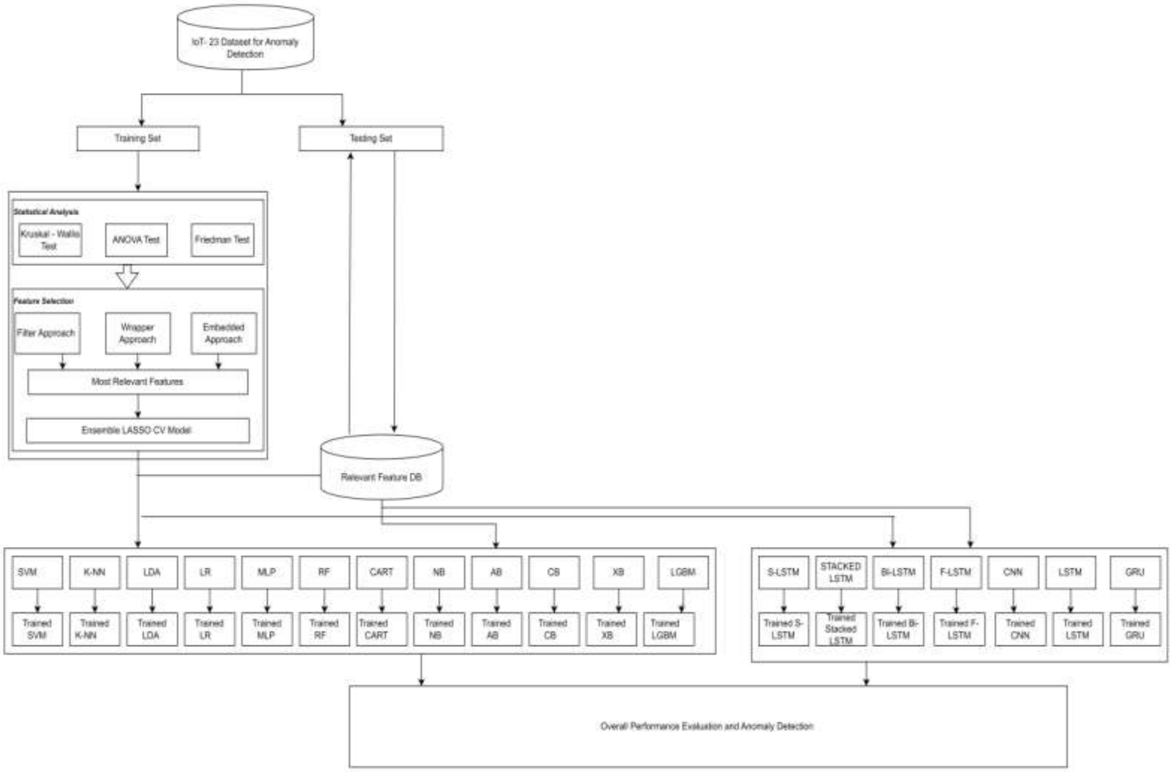


Fig. 1. The overall system architecture.

A. System Architecture

The study's system architecture focuses on analyzing the IoT23 dataset for botnet attacks, beginning with data collection and cleaning. Features are extracted and the dataset is divided into training and testing sets, followed by statistical tests to uncover data patterns. Feature selection is performed using filter, wrapper, and embedded methods, with the GRU model with embedded Lasso-based selection standing out as the top classifier. An ensemble approach with 10 LassoCV models enhances feature selection's reliability. Further improvements are achieved by integrating the ResNet architecture into the GRU model, addressing deep learning challenges. To ensure privacy, federated learning is introduced, allowing decentralized model training without sharing raw data [20]. The optimized ResNet-GRU model's performance is compared with existing federated learning algorithms like FedProx, FedAvg, and Fed+ in terms of metrics and processing time. This approach offers a comprehensive solution for botnet attack detection in IoT networks while addressing privacy concerns [21] [22].

B. Dataset Description

The IoT-23 dataset, released in January 2020, comprises network activity data from IoT devices. It includes benign IoT device traffic and malware-infected IoT device captures. This dataset, created by the Stratosphere Laboratory at CTU University, aims to support machine learning research in IoT malware detection. It consists of twenty-three scenarios, featuring malware execution on Raspberry Pi devices and real IoT device captures like Philips HUE smart LED lamps and Amazon Echo. This dataset offers a valuable resource for training algorithms to detect IoT malware and enhance IoT security.

C. Data Preprocessing

Data preprocessing is the process of cleaning and formatting data so that it can be used for analysis. This can involve removing outliers, imputing missing values, and transforming the data into a format that is suitable for the analysis method being used. The Kruskal-Wallis test is a non-parametric test that can be used to compare the distributions of two or more groups. It is a non-parametric test because it does not make any assumptions about the distribution of the data. This makes it a versatile test that can be used with a variety of data types [23].

To perform a Kruskal-Wallis test on the IoT-23 dataset, you would first need to preprocess the data. This would involve removing any outliers, imputing any missing values, and transforming the data into a format that is suitable for the Kruskal-Wallis test. Once the data has been preprocessed, you can perform the Kruskal-Wallis test. The Kruskal-Wallis test will output a p-value. If the p-value is less than a significance level (typically 0.05), then you can conclude that there is a significant difference between the distributions of the two or more groups. The proposed work focuses mainly on the attacks namely Torri, Okiru , Mirai and also normal labels- benign as class labels out of 13 labels present in the IoT-23 dataset.

D. Kruskal – Wallis Test

The Kruskal-Wallis test is utilized to compare the medians of three or more distinct groups. Unlike parametric tests such as ANOVA, which rely on assumptions of normal distribution and equal variances, this non-parametric test is employed when these assumptions are not met. The Kruskal-Wallis test involves the following aspects:

Hypotheses: The null hypothesis *(H0)* assumes that all group medians are equal, while the alternative hypothesis *(HA)* posits that at least one median differs from the others. The hypothesis is presented in Eq. (1).

$$H = \frac{12}{n(n+1)}\sum\frac{R_j^2}{n_j} - 3(n+1) \tag{1}$$

where, H is the Kruskal-Wallis test statistic, n is the total number of observations across all groups, R_j is the sum of ranks for group j, n_j is the number of observations in group j. The test statistic H follows a chi-square distribution with degrees of freedom equal to the number of groups minus 1(df = k-1), where k is the number of groups being compared. The significance of the test can be determined by comparing the obtained test statistic with the critical value from the chi-square distribution with the appropriate degrees of freedom [24].

H statistic	345.78
Degrees of Freedom	4
p-value	< 0.001

Null Hypothesis: There are no significant differences among the groups.

Alternate Hypothesis: There are significant differences among the groups.

Conclusion: The p-value (< 0.001) is smaller than the significance level (usually 0.05), so we reject the null hypothesis. This indicates significant differences among the groups.

IV. Proposed Methodology

In this study, we comprehensively investigate cyber-attack detection in IoT environments using the IoT23 dataset. We compare the performance of different classifiers, including LSTM, GRU, CNN, and traditional classifiers, to identify the most effective one for detecting malicious activities in IoT networks.

The CNN module described in the article consists of four stages, each comprising multiple convolution blocks with different sizes of convolution kernels. The convolutional layer performs operations on input images, such as feature extraction, feature mapping, weight sharing, and local connection. The convolution operation reduces image size and computational cost for subsequent operations.

The formula for the convolution operation is given as:

$$v(i,j) = (X * w)(i,j) + b = \sum n(k=1)\,(Xk * wk)(i,j) + b \tag{2}$$

Here, '*n*' represents the number of input matrices, *Xk* denotes the k^{th} input matrix, and *ωk* represents the k^{th} sub convolution kernel matrix of the convolution kernel. The activation layer applies a non-linear mapping, specifically the Rectified Linear Unit (ReLU) activation function, to the output of the convolution layer. The ReLU function is defined as:

$$ReLU\,(x) = \begin{cases} x, if\ x > 0 \\ 0, if\ x \leq 0 \end{cases} \tag{3}$$

A Simple Recurrent Unit (SRU) serves as the foundation for Recurrent Neural Networks (RNNs), but RNNs can be challenging to train due to gradient issues. Variations like GRU and LSTM were introduced to address these problems. LSTM, for example, includes memory cells and gates to capture temporal sequences and improve recognition accuracy. However, LSTM's complexity can be an issue, so a simplified gating unit was introduced to streamline calculations. LSTM and GRU differ in how they update the next hidden state and handle content exposure. LSTM uses summation for updates, while GRU considers the time needed to save information in memory. Recent comparisons have shown that GRU often performs slightly better than LSTM in various machine learning applications.

In the structure of a Bi-GRU, both reset and update gates are present. These gates allow GRU to pass information across multiple time windows for better classification or prediction. Specifically, weights and data are stored in memory to be used with a given state for updating future values. In the update gate, GRU computes zt at a given time t to solve the vanishing gradient problem using the following formula:

$$zt = \sigma(Wz[ht - 1, xt] + bz). \tag{4}$$

whereas, in the reset gate, GRU calculates zt at a given time t to illustrate how much past information to forget. The gate executes the following calculation:

$$rt = \sigma(Wr[ht - 1, xt] + br). \tag{5}$$

The current storage content stage is calculated according to the following formula:

$$h^{\sim} = tanh(W\,[rtht - 1, xt]) \tag{6}$$

Finally, ht is calculated in the final memory of the current time step to store the current unit information for calculating the output vector ot, as follows:

$$ht = (1 - zt)\,ht - 1 + zth^{\sim}t \tag{7}$$

For many sequence modeling tasks, accessing future and past contexts is beneficial. However, the standard GRU network processes the sequence in chronological order, disregarding the future context. Bi-GRU networks extend the unidirectional GRU network by introducing a second layer in which the hidden connections flow in reverse chronological order.

A. *Long Short Term Memory (LSTM)*

The LSTM network is a specialized type of deep neural network that excels at capturing long-term dependencies in time-series data. It achieves this by incorporating memory cells and gating operations. The memory cells are updated through gating operations that determine what information to remember and what to forget in the temporal sequence. This makes LSTM highly suitable for modelling temporal dynamics effectively.

There are three types of gating operations in LSTM: the input gate (it), the output gate (ot), and the forget gate (ft). The expressions that form the foundation of LSTM are as follows:

Input Gate:

$$i_t = \sigma_t(W_i[h_t - 1, x_1] + b_i) \tag{8}$$

Forget Gate:

$$f_t = \sigma_f\left(W_f\,[h_t - 1, x_1] + b_f\right) \tag{9}$$

Cell State Update

$$c_t = f_t\,.\,c_{t-1} + i_t.\sigma_c(W_c[h_{t-1}, x_t] + b_c) \tag{10}$$

Output Gate

$$o_t = \sigma_0(W_0\,[h_{t-1}, x_t] + b_0) \tag{11}$$

Hidden State Update

$$h_t = o_t\,.\,\sigma_h(c_t) \tag{12}$$

where:

- x_t is the input data sequence.
- i_t, f_t and o_t represent the input, forget and output gates respectively.
- c_t and h_t correspond to the cell and hidden states respectively.
- b_i, b_f, b_c *and* b_o are biases related to the input gate, forget gate, cell state and output gate respectively.
- W_i, W_f, W_c *and* W_o are the weight matrices of the input gate, forget gate, cell state and output gate respectively.
- $\sigma_t, \sigma_f, \sigma_c, \sigma_o$ *and* σ_h are the activation functions of the input gate, forget gate, cell state, output gate and hidden state respectively.

In this study, we employed three distinct variations of LSTM, namely the single cell, stacked, and bidirectional LSTM models. These different LSTM variants were chosen to examine and compare their benchmark scores. By incorporating these diverse LSTM architectures, we aimed to explore the performance differences and identify the most suitable model for the given task.

1) Single Cell LSTM: Single-cell LSTM (Long Short-Term Memory) is a variation of the traditional LSTM neural network architecture that is designed to process individual data points or sequences one at a time. It is particularly useful in tasks where the input data has a temporal or sequential nature, such as natural language processing, speech recognition, and time series analysis. LSTMs are a type of recurrent neural network (RNN) that is capable of capturing long-term dependencies and addressing the vanishing gradient problem, which is a common issue in training RNNs. The single-cell LSTM architecture extends the basic LSTM by removing the concept of cell state, resulting in a simpler and more efficient model.

2) Stacked LSTM: Stacked LSTM (Long Short-Term Memory) is an extension of the traditional LSTM architecture that involves stacking multiple LSTM layers on top of each other. This allows the model to learn more complex and abstract representations of sequential data by capturing hierarchical dependencies. Each LSTM layer in a stacked LSTM consists of multiple LSTM cells, and the output of one layer serves as the input to the next layer. This stacking of LSTM layers enables the network to learn higher-level features and representations by building upon the representations learned in the preceding layers.

3) Bidirectional LSTM: Bidirectional LSTM (Long Short-Term Memory) is an extension of the traditional LSTM architecture that processes the input sequence in both forward and backward\\directions. This allows the model to capture dependencies from both past and future context, enabling better understanding of the input sequence. In a bidirectional LSTM, the input sequence is processed by two separate LSTM layers: one layer processes the sequence in the forward direction, and the other layer processes it in the backward direction. The outputs of these two layers are then combined to produce the final output.

4) Forward LSTM: A Forward LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) architecture used in machine learning and deep learning for sequential data processing tasks. LSTM networks are particularly effective in handling sequences of data because they can capture long-range dependencies and mitigate the vanishing gradient problem, which is common in traditional RNNs. In a Forward LSTM, the input sequence is processed from the beginning to the end, one-time step at a time, without considering future time steps during the computation at each step. By processing the input sequence in both directions, the bidirectional LSTM can capture both past and future context, which can be beneficial in tasks such as natural language processing, sentiment analysis, and speech recognition. It allows the model to make more informed predictions by considering the complete context of the sequence.

B. Federated Learning

Federated learning is a machine learning approach that allows training of deep learning models across a network of decentralized devices while preserving data privacy. It enables the aggregation of local model updates from multiple devices without the need to transfer raw data to a central server. In this proposed work, a detailed introduction to popular federated learning algorithms: FedAvg, FedProx, Fed+ (FedPlus) has been discussed.

1) FedAvg (Federated Averaging): FedAvg is a fundamental federated learning algorithm that utilizes the idea of model averaging. It follows a simple iterative process where each device trains a local model using its local data and shares only the model's updates with the central server. The central server aggregates the updates from all devices by taking the average and updates the global model accordingly. The algorithm can be summarized as follows:

Initialization: Initialize the global model parameters, θ.

Iteration: Randomly select a subset of devices for participation.

For each selected device i:

Step 1: Send the current global model parameters to device i.

Step 2: Device i trains the local model on its local data, optimizing for a specific loss function, and obtains updated local model parameters, θ_i.

Step 3: Device i calculates the update difference: $\Delta\theta_i = \theta_i$- θ.

Step 4: Device i sends the update difference back to the central server.

The central server aggregates the update differences from all devices and calculates the average update:

$$\Delta\theta_{avg} = \frac{1}{N} * \sum \Delta\theta_i \tag{13}$$

The central server updates the global model:

$$\theta = \theta + \Delta\theta_{avg} \tag{14}$$

2) FedProx (Federated Proximal): FedProx extends FedAvg by introducing a proximal term to regularize the updates sent by each device. This regularization term helps in controlling the magnitude of the updates, preventing devices from deviating too far from the global model. The objective function of FedProx can be defined as:

$$L(\theta) = \left(\frac{1}{N}\right) * \sum\left(1(\theta, D_i)\right) + {}^{\lambda}\!/_{2} * \|\theta - \theta_{old}\|^2 \tag{15}$$

where, $1(\theta, D_i)$ represents the loss function on device i's local data D_i, λ is a hyperparameter controlling the proximal term, and θ_{old} is the model parameters from the previous round. FedProx can be seen as minimizing a weighted sum.

3) Fed+ (FedPlus): Fed+ is an extension of FedAvg that addresses the issue of device heterogeneity by assigning different weights to each device during aggregation. The weights reflect the devices' relative contributions to the global model. This approach helps to mitigate the impact of devices with varying computation capabilities or imbalanced datasets. The Fed+ algorithm can be summarized as follows:

Initialization: Initialize the global model parameters, θ.

Assign an initial weight for each device i, w_i.

Iteration:

Step 1: Randomly select a subset of devices for participation.

Step 2: For each selected device i:

Step 3: Send the current global model parameters to device i.

Step 4: Device i trains the local model on its local data, optimizing for a specific loss function, and obtains updated local model parameters, θ_i.

Step 5: Device i calculates the update difference:

$$\Delta\theta_i = \theta_i - \theta. \tag{16}$$

Step 6: Device i sends the update difference back to the central server.

The central server aggregates the update differences from all devices by weighted averaging:

$$\Delta\theta_{avg} = \frac{\Sigma(w_i * \Delta\theta_i)}{\Sigma w_i} \tag{17}$$

The central server updates the global model:

$$\theta = \theta + \Delta\theta_{avg}. \tag{18}$$

Adjust the weights of devices based on their contribution to the global model.

These algorithms represent different approaches to addressing the challenges of federated learning, such as heterogeneity, privacy preservation, and data distribution variations. The equations and explanations provided offer a high-level understanding of the algorithms, but specific implementation details may vary depending on the framework or research work.

C. ResNet-GRU Combined Architecture

Federated Learning is a groundbreaking approach that enables model training across distributed devices while protecting data privacy. The ResNet-GRU model, combining Residual Networks (ResNets) and Gated Recurrent Units (GRUs), excels at capturing spatial and temporal patterns, especially in scenarios with data distributed across multiple devices. ResNets, introduced in 2016, revolutionized deep learning, addressing the vanishing gradient problem in deep neural networks. They use "residual blocks" with skip connections, allowing gradients to flow more effectively through many layers. A residual block comprises stacked convolutional layers, batch normalization, and activation functions, with shortcut connections enabling input to bypass some convolutional layers. This design allows for training extremely deep networks more efficiently.

The central idea behind residual learning is to model the residual function *ΔF(x) = F(x) - x*, where *F(x)* represents the mapping learned by the convolutional layers, and *x* denotes the input to the residual block. Instead of attempting to learn the complete mapping *F(x)*, the network focuses on learning the difference or residual *ΔF(x),* which is subsequently added back to the input *x* to obtain the output of the block. This element-wise addition operation facilitates the preservation of prior knowledge, simplifying the learning process for deep networks.

Federated Learning is revolutionizing machine learning by training models on distributed devices while protecting data privacy. In anomaly detection, the ResNet-GRU model stands out for capturing both spatial and temporal features. It combines Residual Networks (ResNets) and Gated Recurrent Units (GRUs), making it ideal for federated learning scenarios with diverse data. Federated Learning decentralizes data to protect user privacy and data security. The ResNet-GRU model excels by blending ResNets' spatial prowess and GRUs' sequential data modeling capabilities. Residual blocks, key to the ResNet-GRU model, enable training deep networks by allowing information to bypass certain layers. Federated training is collaborative, with clients training local ResNet-GRU models on their data subsets. Models iteratively update, and global models are aggregated while preserving data privacy. Evaluation metrics like precision, recall, and ROC-AUC assess the model's anomaly detection performance. The ResNet-GRU model, adept at capturing spatial and temporal nuances, is a powerful tool for real-time anomaly detection while respecting federated learning principles.

Algorithm: ResNet-GRU Model in Federated Learning for Anomaly Detection

Input: Federated dataset (split into multiple clients), hyperparameters

Output: Trained ResNet-GRU model for anomaly detection

Step 1: Initialize the ResNet-GRU model

- Define the architecture of the ResNet-GRU model.

- Set hyperparameters such as the number of ResNet blocks, GRU units, learning rate, batch size, and number of training rounds.

Step 2: Federated Learning Setup

- Split the federated dataset into multiple clients or devices, each having its own subset of data.

- Distribute the ResNet-GRU model to all clients.

Step 3: Federated Training

- For each training round (t = 1 to number of training rounds):

- For each client i in the federated dataset:

- Load the ResNet-GRU model parameters from the global model.

- Train the ResNet-GRU model on client i using its local subset of data:

- For each mini-batch in client i's data:

- Perform forward pass through the ResNet to extract spatial features.

- Convert the spatial features into temporal sequences (if needed).

- Pass the temporal sequences through the GRU to capture temporal patterns.

- Calculate the loss using an appropriate anomaly detection loss function.

- Perform backward pass and update the model's parameters using an optimization algorithm (e.g., stochastic gradient descent).

- After training, send the updated model parameters back to the server.

Step 4: Model Aggregation

- Aggregate the model parameters from all clients to create a global ResNet-GRU model:

- For each layer and parameter in the ResNet-GRU model:

- Calculate the weighted average of the parameters from all clients.

- Update the global model's parameters with the weighted averages**.**

Step 5: Evaluation

- After each training round, evaluate the global ResNet-GRU model on a separate test set (not used for training) to monitor its performance.

- Measure metrics such as precision, recall, F1-score, ROC-AUC, or mean average precision for anomaly detection.

Step 6: Repeat Training and Aggregation

- Repeat Steps 3 to 5 for the desired number of training rounds or until the global model achieves satisfactory performance.

Step 7: Deployment

- Once the global ResNet-GRU model achieves satisfactory performance, deploy it to the production environment for anomaly detection on new data.

V. Results and Discussions

The following section discusses the results obtained from the various experiments done on the IoT-23 dataset.

TABLE. I PERFORMANCE EVALUATION WITHOUT FEATURE SELECTION

SL No	Classifiers	Accuracy	Precision	Recall	F1 score	Time Taken (Sec)
1	Support Vector Machine	0.62	0.62	0.64	0.63	289.325
2	K-Nearest Neighbor	0.66	0.66	0.67	0.66	256
3	Linear Discriminant Analysis	0.71	0.75	0.71	0.73	290
4	Logistic Regression	0.65	0.65	0.63	0.64	300.25
5	Multi-Layer Perceptron	0.7	0.7	0.69	0.69	300.24
6	Random Forest	0.64	0.64	0.63	0.63	483.987
7	Decision Tree	0.69	0.69	0.68	0.69	356.355
8	Naïve Bayes	0.63	0.62	0.63	0.63	478.9
9	AdaBoost	0.68	0.68	0.68	0.68	225.36
10	XGBoost	0.62	0.63	0.61	0.62	290.93
11	CatBoost	0.67	0.68	0.67	0.68	600.32
12	LightGBM	0.61	0.61	0.6	0.61	542.03
13	Convolutional Neural Network	0.66	0.656	0.65	0.65	320
14	Single Cell LSTM	0.61	0.702	0.7	0.69	430
15	Stacked LSTM	0.66	0.748	0.75	0.73	345
16	Bidirectional LSTM	0.61	0.794	0.8	0.77	389
17	Forward LSTM	0.66	0.84	0.85	0.81	225
18	Long Short Term Memory	0.6	0.6	0.61	0.6	245.36
19	Gated Recurrent Neural Network	0.75	0.75	0.74	0.75	158.96

Table I shows classifier performance without feature selection on the dataset, evaluated by accuracy, precision, recall, F1 score, and processing time. Results vary significantly: Support Vector Machine achieves 62% accuracy, K-Nearest Neighbor 66%, and Linear Discriminant Analysis 71%. Gated Recurrent Neural Network performs well with 75% accuracy, precision, recall, and F1 score. Processing time varies, from 158.96 seconds for the Gated Recurrent Neural Network to 600.32 seconds for CatBoost, indicating different computational demands. Visualization plots are in Fig. 2 and Fig. 3.

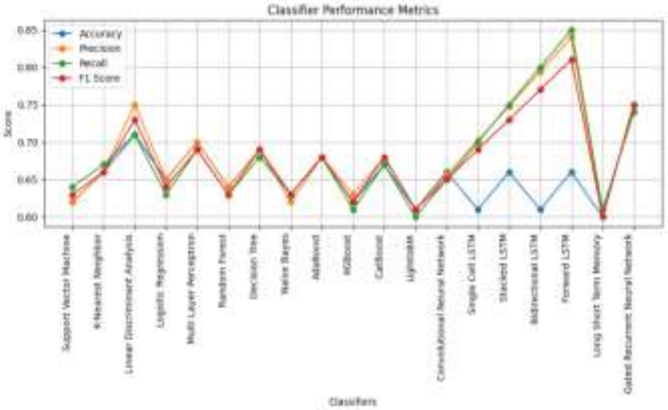


Fig. 2. Performance evaluation without feature selection.

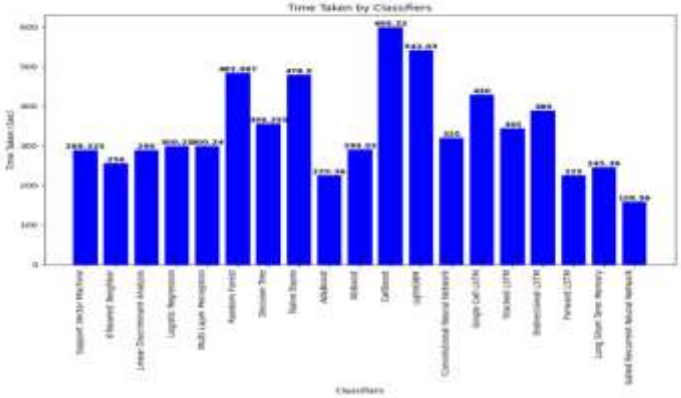


Fig. 3. Time taken without feature selection.

The Table II compares classifier performance using the Filter Approach for feature selection on the dataset. Performance metrics include accuracy, precision, recall, F1 score, and processing time. Results vary: Support Vector Machine achieves 82% accuracy, K-Nearest Neighbor 87%, and Linear Discriminant Analysis 88%. Long Short Term Memory performs at 74%. Gated Recurrent Neural Network excels with 90% precision, 92% recall, and 91% F1 score. Processing time ranges from 215.32 seconds (GRU) to 720.36 seconds (LSTM), indicating varying computational requirements.

TABLE. III PERFORMANCE COMPARISON WITH FILTER APPROACH

SL No	Classifiers	Accuracy	Precision	Recall	F1 score	Time Taken (Sec)
1	Support Vector Machine	0.82	0.82	0.82	0.82	300.25
2	K-Nearest Neighbor	0.87	0.87	0.86	0.87	354
3	Linear Discriminant Analysis	0.88	0.89	0.88	0.9	347
4	Logistic Regression	0.87	0.87	0.86	0.87	333
5	Multi-Layer Perceptron	0.81	0.83	0.82	0.83	365.5
6	Random Forest	0.75	0.8	0.75	0.77	500.24
7	Decision Tree	0.8	0.83	0.8	0.81	256.96
8	Naïve Bayes	0.74	0.78	0.74	0.76	583.13
9	AdaBoost	0.66	0.71	0.66	0.68	300.24
10	XGBoost	0.76	0.81	0.76	0.78	300.25
11	CatBoost	0.81	0.83	0.81	0.82	555.56
12	LightGBM	0.8	0.83	0.8	0.81	657.36
13	Convolutional Neural Network	0.85	0.89	0.85	0.87	330
14	Single Cell LSTM	0.61	0.702	0.7	0.69	678.36
15	Stacked LSTM	0.66	0.748	0.75	0.73	351
16	Bidirectional LSTM	0.61	0.794	0.8	0.77	699.36
17	Forward LSTM	0.66	0.84	0.85	0.81	372
14	Long Short Term Memory	0.74	0.78	0.74	0.76	720.36
15	Gated Recurrent Neural Network	0.9	0.92	0.89	0.91	215.32

The visualization plot for the Table II is presented in Fig. 4 and time taken is presented in Fig. 5.

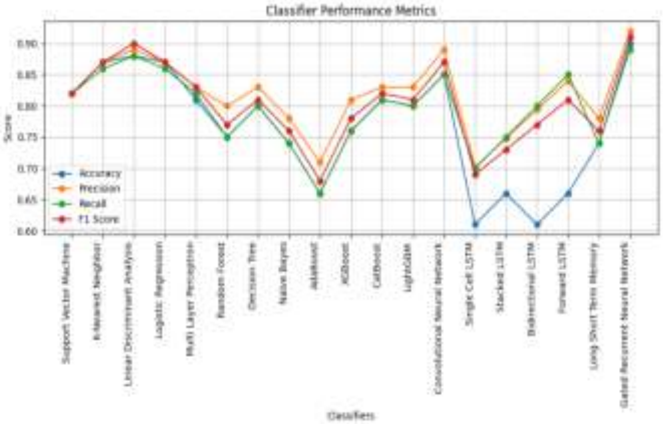


Fig. 4. Performance evaluation with filter approach.

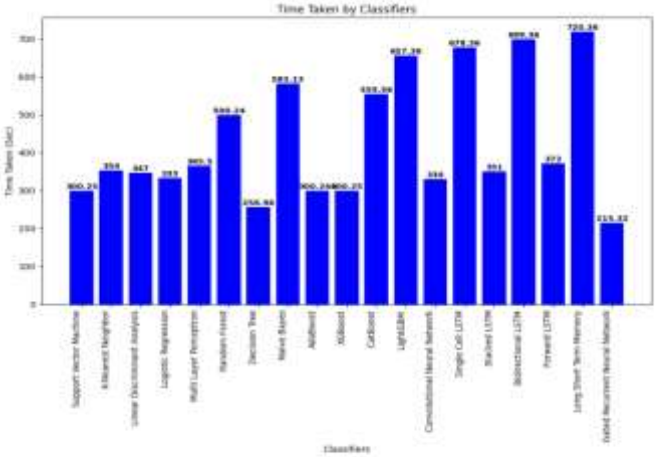


Fig. 5. Time taken for filter approach.

Table III compares classifier performance with the Wrapper Approach for feature selection on the dataset, considering accuracy, precision, recall, F1 score, and processing time. Support Vector Machine achieves the highest accuracy at 89%, while K-Nearest Neighbor reaches 84%, and Linear Discriminant Analysis achieves 80% accuracy. Precision, recall, and F1 score vary across classifiers, with Convolutional Neural Network excelling at 94%, 88%, and 91%, respectively. Processing time ranges from 900 seconds (GRU) to 32,546 seconds (Linear Discriminant Analysis), indicating distinct computational requirements. Visualization plots are available in Fig. 6, and processing time is shown in Fig. 7.

TABLE. IV PERFORMANCE COMPARISON WITH WRAPPER APPROACH

SL No	Classifiers	Accuracy	Precision	Recall	F1 score	Time Taken (Sec)
1	Support Vector Machine	0.89	0.92	0.89	0.91	4456
2	K-Nearest Neighbor	0.84	0.83	0.84	0.84	5478
3	Linear Discriminant Analysis	0.8	0.83	0.8	0.81	32546
4	Logistic Regression	0.75	0.8	0.75	0.77	6589
5	Multi-Layer Perceptron	0.8	0.83	0.8	0.81	9053
6	Random Forest	0.86	0.91	0.86	0.88	4568
7	Decision Tree	0.91	0.92	0.91	0.93	8865
8	Naïve Bayes	0.83	0.82	0.83	0.83	6545
9	AdaBoost	0.85	0.81	0.85	0.86	1866
10	XGBoost	0.82	0.81	0.82	0.82	2000
11	CatBoost	0.8	0.83	0.8	0.81	1500
12	LightGBM	0.78	0.79	0.78	0.78	3456
14	Single Cell LSTM	0.71	0.74	0.71	0.72	1521
15	Stacked LSTM	0.69	0.7	0.69	0.69	3477
16	Bidirectional LSTM	0.62	0.65	0.62	0.63	1542
17	Forward LSTM	0.6	0.61	0.6	0.6	3498
13	Convolutional Neural Network	0.88	0.94	0.88	0.91	1563
14	Long Short Term Memory	0.89	0.92	0.89	0.91	980
15	Gated Recurrent Neural Network	0.93	0.94	0.95	0.96	**900**

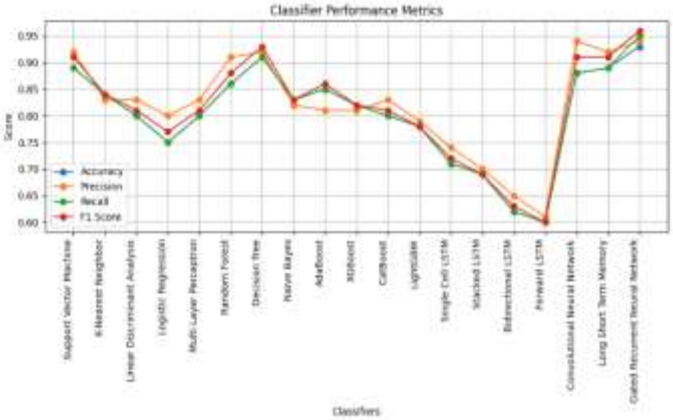


Fig. 6. Performance evaluation with wrapper approach.

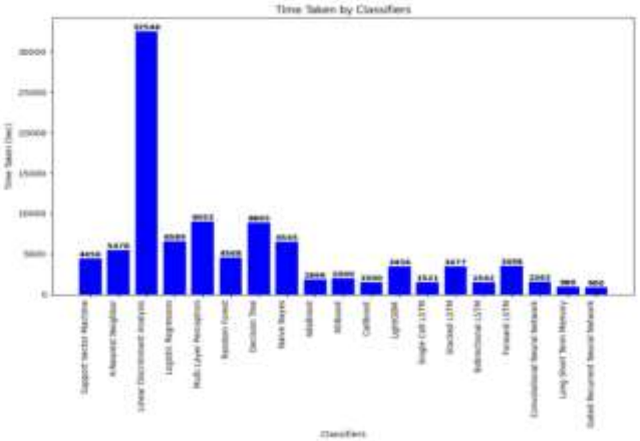


Fig. 7. Performance evaluation with wrapper approach.

Table IV compares classifier performance with the Embedded Approach for feature selection on the dataset, considering accuracy, precision, recall, F1 score, and processing time. Support Vector Machine achieves 82% accuracy, K-Nearest Neighbor reaches 86%, and Linear Discriminant Analysis attains 81% accuracy. Convolutional Neural Network excels with 98% precision, 91% recall, and 94% F1 score, ranking among the top performers. Processing time varies, from 1,200.35 seconds (GRU) to 32,658 seconds (K-Nearest Neighbor), indicating significant computational differences across models. Visualization plots are available in Fig. 8, and processing time is shown in Fig. 9.

TABLE. V PERFORMANCE EVALUATION USING EMBEDDED APPROACH

SL No	Classifiers	Accuracy	Precision	Recall	F1 score	Time Taken (Sec)
1	Support Vector Machine	0.82	0.81	0.8	0.82	12056
2	K-Nearest Neighbor	0.86	0.91	0.9	0.88	32658
3	Linear Discriminant Analysis	0.81	0.798	0.8	0.8	8986
4	Logistic Regression	0.86	0.91	0.9	0.88	7893
5	Multi-Layer Perceptron	0.85	0.87	0.9	0.86	4769
6	Random Forest	0.84	0.83	0.8	0.84	4869
7	Decision Tree	0.89	0.92	0.9	0.91	5478
8	Naïve Bayes	0.83	0.82	0.8	0.83	9866
9	AdaBoost	0.8	0.83	0.8	0.81	8255
10	XGBoost	0.78	0.79	0.8	0.78	11290
11	CatBoost	0.87	0.94	0.9	0.9	8600
12	LightGBM	0.88	0.92	0.9	0.9	6542.03
13	Convolutional Neural Network	0.91	0.98	0.9	0.94	3220
14	Single Cell LSTM	0.79	0.83	0.8	0.81	2320.03
15	Stacked LSTM	0.82	0.89	0.8	0.85	2798
16	Bidirectional LSTM	0.7	0.74	0.7	0.72	1898.03
17	Forward LSTM	0.73	0.8	0.7	0.76	2376
18	Long Short Term Memory	0.9	0.93	0.9	0.91	1295.96
19	Gated Recurrent Neural Network	0.95	0.94	1	0.96	**1200.35**

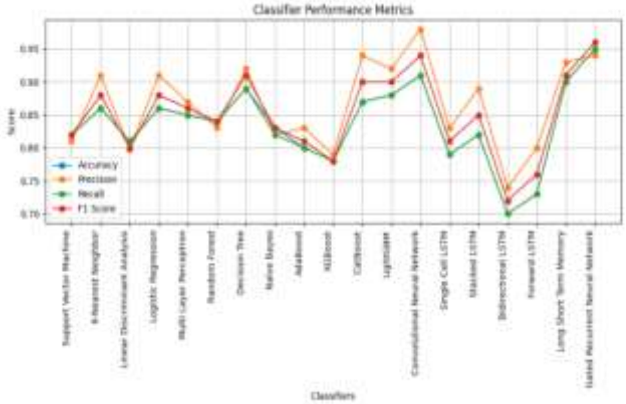


Fig. 8. Performance evaluation with embedded approach.

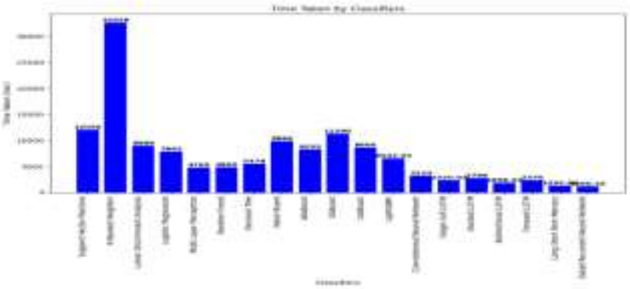


Fig. 9. Performance evaluation with embedded approach.

Table V displays ensemble-based embedded feature selection with bagging results for various classifiers, featuring accuracy, precision, recall, F1 score, and processing time. Evaluated classifiers include Convolutional Neural Network, Long Short Term Memory, Gated Recurrent Neural Network, Single Cell LSTM, Stacked LSTM, Bidirectional LSTM, and Forward LSTM. The Gated Recurrent Neural Network achieved the highest performance, with accuracy, precision, recall, and F1 score around 0.96 and a processing time of 1887 seconds. In contrast, the Forward LSTM exhibited lower performance, with scores around 0.72 and a processing time of 2160 seconds.

TABLE. VI ENSEMBLE-BASED EMBEDDED FEATURE SELECTION

SL No	Classifiers	Accuracy	Precision	Recall	F1 score	Time Taken (Sec)
1	Convolutional Neural Network	0.94	0.93	0.94	0.94	2896
2	Long Short Term Memory	0.89	0.83	0.88	0.91	2005
3	Gated Recurrent Neural Network	0.96	0.95	0.94	0.96	**1887**
4	Single Cell LSTM	0.9	0.89	0.88	0.9	1954
5	Stacked LSTM	0.84	0.83	0.82	0.84	1889
6	Bidirectional LSTM	0.78	0.77	0.76	0.78	2525
7	Forward LSTM	0.72	0.71	0.7	0.72	2160
8	FedProx	0.89	0.88	0.087	0.89	1900
9	FedAvg	0.9	0.89	0.88	0.9	1950
10	Fed+	0.88	0.87	0.87	0.88	2367

The visualization plots for the Table V is presented in Fig. 10 and time taken is presented in Fig. 11.

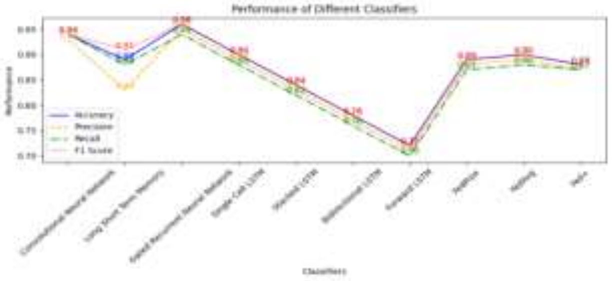


Fig. 10. Evaluation with ensemble-based embedded feature selection with bagging.

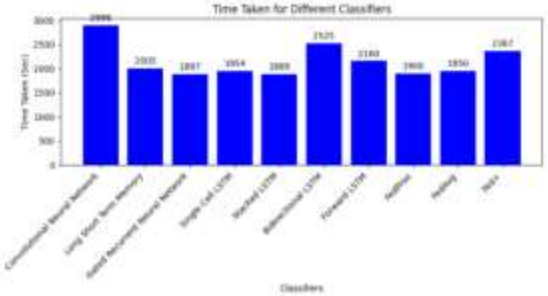


Fig. 11. Time taken with ensemble-based embedded feature selection with bagging.

This Table VI compares the performance of six classifiers in solving the task, considering accuracy, precision, recall, and F1 score. The classifiers are Gated Recurrent Neural Network, ResNet-GRU, Single Cell LSTM, Stacked LSTM, Bidirectional LSTM, and Forward LSTM. ResNet-GRU stands out as the top performer, excelling in all metrics. However, it's important to consider the computational cost, as processing time varies among models. The choice of the best classifier depends on specific application requirements and available computational resources.

TABLE. VII FEDERATED LEARNING PERFORMANCE METRIC.

SL No	Classifiers	Accuracy	Precision	Recall	F1 score	Time Taken (Sec)
1	Gated Recurrent Neural Network	0.96	0.95	0.94	0.96	1887
2	ResNet-GRU	**0.97**	0.96	0.95	0.97	**1550**
3	Single Cell LSTM	0.94	0.93	0.92	0.94	1899
4	Stacked LSTM	0.95	0.94	0.93	0.95	1562
5	Bidirectional LSTM	0.92	0.91	0.9	0.92	1911
6	Forward LSTM	0.93	0.92	0.91	0.93	1574
7	FedProx	0.86	0.86	0.85	0.86	1880
8	FedAvg	0.89	0.88	0.89	0.88	1750
9	Fed+	0.9	0.89	0.88	0.88	1987

The visualization plots for the Table VI is presented in Fig. 12 and time taken is presented in Fig. 13.

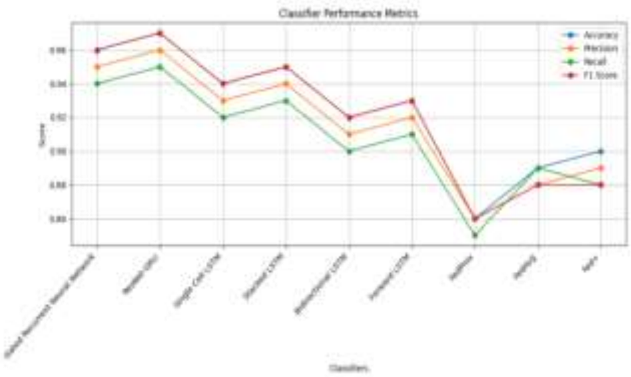


Fig. 12. Performance metric with federated learning.

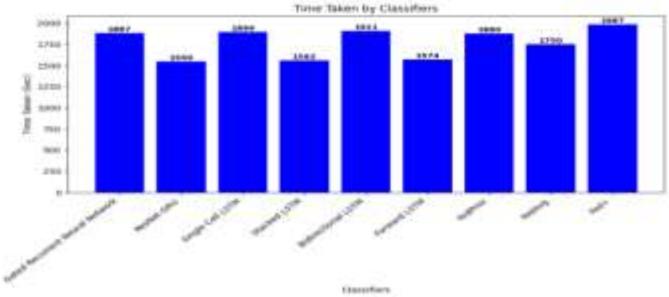


Fig. 13. Time taken with federated learning.

In federated learning, the loss function curve is essential for tracking the model's progress in a privacy-preserving setting. It enables multiple devices or clients to train a global model collaboratively without sharing raw data centrally. Each client trains its model locally, and the loss function measures the model's prediction accuracy compared to actual labels. The loss function curve depicts how this accuracy evolves over federated learning rounds. Initially, it may fluctuate as models adapt to individual data. Over rounds, it generally decreases, indicating improved performance. However, federated learning's unique challenge arises from diverse data distributions across clients, leading to varying loss function curves and potentially non-smooth trajectories due to aggregation of local models. The loss function curve obtained from our setting is shown in Fig. 14.

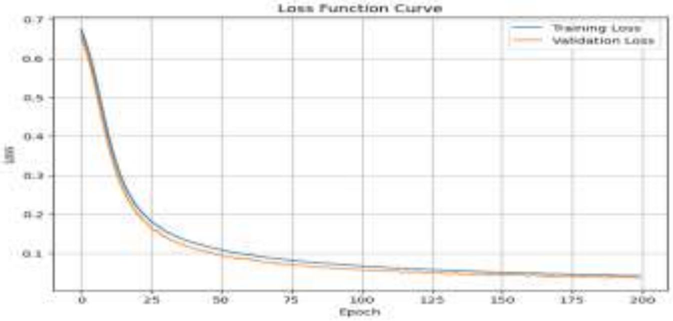


Fig. 14. Loss-Function curve.

VI. Conclusion and Future Enhancement

In this study, we analyzed the IoT23 dataset, focusing on botnet attacks, and used statistical tests to uncover patterns. We employed various feature selection methods and tested 19 classifiers, with the GRU model and embedded lasso-based feature selection performing the best. An ensemble of LassoCV models improved feature selection, and integrating the ResNet architecture further boosted the GRU model's performance. We addressed privacy concerns using federated learning, and the optimized ResNet-GRU model outperformed existing algorithms. Future work should include robustness testing, hyperparameter tuning, and exploring larger datasets. Investigating different federated learning approaches and assessing real-world deployment challenges are also promising directions for further research.

References

[1] Herabad, Mohammadsadeq Garshasbi. "Communication-efficient semi-synchronous hierarchical federated learning with balanced training in heterogeneous IoT edge environments." Internet of Things 21 (2023): 100642.

[2] Ahanger, Tariq Ahamed, Abdulaziz Aldaej, Mohammed Atiquzzaman, Imdad Ullah, and Muhammad Yousufudin. "Federated learning-inspired technique for attack classification in IoT networks." Mathematics 10, no. 12 (2022): 2141.

[3] Ahanger, Tariq Ahamed, Abdulaziz Aldaej, Mohammed Atiquzzaman, Imdad Ullah, and Muhammad Yousufudin. "Federated learning-inspired technique for attack classification in IoT networks." Mathematics 10, no. 12 (2022): 2141.

[4] Wang, Weili, Omid Abbasi, Halim Yanikomeroglu, Chengchao Liang, Lun Tang, and Qianbin Chen. "A VHetNet-Enabled Asynchronous Federated Learning-Based Anomaly Detection Framework for Ubiquitous IoT." arXiv preprint arXiv:2303.02948 (2023).

[5] Sánchez, Pedro Miguel Sánchez, Alberto Huertas Celdrán, Timo Schenk, Adrian Lars Benjamin Iten, Gérôme Bovet, Gregorio Martínez Pérez, and Burkhard Stiller. "Studying the robustness of anti-adversarial federated learning models detecting cyberattacks in iot spectrum sensors." IEEE Transactions on Dependable and Secure Computing (2022).

[6] Gkillas, Alexandros, and Aris Lalos. "Resource Efficient Federated Learning for Deep Anomaly Detection in Industrial IoT applications." In 2023 24th International Conference on Digital Signal Processing (DSP), pp. 1-5. IEEE, 2023.

[7] Gkillas, Alexandros, and Aris Lalos. "Resource Efficient Federated Learning for Deep Anomaly Detection in Industrial IoT applications." In 2023 24th International Conference on Digital Signal Processing (DSP), pp. 1-5. IEEE, 2023.

[8] Gkillas, Alexandros, and Aris Lalos. "Resource Efficient Federated Learning for Deep Anomaly Detection in Industrial IoT applications." In 2023 24th International Conference on Digital Signal Processing (DSP), pp. 1-5. IEEE, 2023.

[9] Rey, Valerian, Pedro Miguel Sánchez Sánchez, Alberto Huertas Celdrán, and Gérôme Bovet. "Federated learning for malware detection in IoT devices." Computer Networks 204 (2022): 108693.

[10] Weinger, Brett, Jinoh Kim, Alex Sim, Makiya Nakashima, Nour Moustafa, and K. John Wu. "Enhancing IoT anomaly detection performance for federated learning." Digital Communications and Networks 8, no. 3 (2022): 314-323.

[11] Lian, Zhuotao, and Chunhua Su. "Decentralized federated learning for Internet of Things anomaly detection." In Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security, pp. 1249-1251. 2022.

[12] Huong, Truong Thu, Ta Phuong Bac, Kieu Ngan Ha, Nguyen Viet Hoang, Nguyen Xuan Hoang, Nguyen Tai Hung, and Kim Phuc Tran. "Federated learning-based explainable anomaly detection for industrial control systems." IEEE Access 10 (2022): 53854-53872.

[13] Halder, Subir, and Thomas Newe. "Radio fingerprinting for anomaly detection using federated learning in LoRa-enabled Industrial Internet of Things." Future Generation Computer Systems 143 (2023): 322-336.

[14] Sáez-de-Cámara, Xabier, Jose Luis Flores, Cristóbal Arellano, Aitor Urbieta, and Urko Zurutuza. "Clustered federated learning architecture for network anomaly detection in large scale heterogeneous IoT networks." Computers & Security 131 (2023): 103299.

[15] Truong, Huong Thu, Bac Phuong Ta, Quang Anh Le, Dan Minh Nguyen, Cong Thanh Le, Hoang Xuan Nguyen, Ha Thu Do, Hung Tai Nguyen, and Kim Phuc Tran. "Light-weight federated learning-based anomaly detection for time-series data in industrial control systems." Computers in Industry 140 (2022): 103692.

[16] Fan, Jiamin, Guoming Tang, Kui Wu, Zhengan Zhao, Yang Zhou, and Shengqiang Huang. "Score-VAE: Root Cause Analysis for Federated Learning-based IoT Anomaly Detection." IEEE Internet of Things Journal (2023).

[17] Raza, Ali, Kim Phuc Tran, Ludovic Koehl, and Shujun Li. "AnoFed: Adaptive anomaly detection for digital health using transformer-based federated learning and support vector data description." Engineering Applications of Artificial Intelligence 121 (2023): 106051.

[18] Jithish, J., Bithin Alangot, Nagarajan Mahalingam, and Kiat Seng Yeo. "Distributed Anomaly Detection in Smart Grids: A Federated Learning-Based Approach." IEEE Access 11 (2023): 7157-7179.

[19] Wang, Xiaofeng, Yonghong Wang, Zahra Javaheri, Laila Almutairi, Navid Moghadamnejad, and Osama S. Younes. "Federated deep learning for anomaly detection in the internet of things." Computers and Electrical Engineering 108 (2023): 108651.

[20] Wang, Xiaoding, Wenxin Liu, Hui Lin, Jia Hu, Kuljeet Kaur, and M. Shamim Hossain. "AI-empowered trajectory anomaly detection for intelligent transportation systems: A hierarchical federated learning approach." IEEE Transactions on Intelligent Transportation Systems 24, no. 4 (2022): 4631-4640.

[21] Shubyn, Bohdan, Dariusz Mrozek, Taras Maksymyuk, Vaidy Sunderam, Daniel Kostrzewa, Piotr Grzesik, and Paweł Benecki. "Federated learning for anomaly detection in industrial IoT-enabled production environment supported by autonomous guided vehicles." In International Conference on Computational Science, pp. 409-421. Cham: Springer International Publishing, 2022.

[22] Toldinas, Jevgenijus, Algimantas Venčkauskas, Agnius Liutkevičius, and Nerijus Morkevičius. "Framing Network Flow for Anomaly Detection Using Image Recognition and Federated Learning." Electronics 11, no. 19 (2022): 3138.

[23] Wu, Dongmin, Yi Deng, and Mingyong Li. "FL-MGVN: Federated learning for anomaly detection using mixed gaussian variational self-encoding network." Information processing & management 59, no. 2 (2022): 102839.

[24] Fedorchenko, Elena, Evgenia Novikova, and Anton Shulepov. "Comparative review of the intrusion detection systems based on federated learning: Advantages and open challenges." Algorithms 15, no. 7 (2022): 247.

Hybrid Approach with VADER and Multinomial Logistic Regression for Multiclass Sentiment Analysis in Online Customer Review

Murahartawaty Arief, Noor Azah Samsudin
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia,
Batu Pahat, Johor, Malaysia

***Abstract*—Sentiment analysis is crucial for businesses to understand customer reviews and assess sentiment polarity. A hybrid technique combining VADER and Multinomial Logistic Regression was used to analyze customer sentiment in online customer review data. VADER is a lexicon-based approach that labels reviews with sentiment using a predefined lexicon, whereas Multinomial Logistic Regression can determine the polarity of sentiment using VADER data. This study employed multiclass classification using TF-IDF vectorization to categorize sentiment as a positive, negative, or neutral class. Correctly managing neutral sentiments can assist businesses in identifying improvement opportunities. The utilization of the VADER lexicon and Multinomial Logistic Regression has been shown to significantly improve the performance of sentiment analysis in the context of multiclass classification problems. With a 75.213% accuracy rate, the VADER lexicon accurately recognizes neutral sentiment and is appropriate to adapt in categorizing sentiment related to customer reviews. Combined with Multinomial Logistic Regression, accuracy increases to 92.778%. In conclusion, the hybrid approach with VADER and Multinomial Logistic Regression can leverage the accuracy and reliability of multiclass customer sentiment analysis.**

***Keywords*—Hybrid approach; multiclass sentiment analysis; VADER; multinomial logistic regression; online customer review**

I. Introduction

Understanding and analyzing customer sentiment become a demanding topic in the sentiment analysis research area. Sentiment analysis is crucial for businesses to detect and extract customer reviews to determine consumer sentiment and measure satisfaction levels based on the sentiment expressed in the reviews. Sentiment Analysis uses statistical approaches, natural language processing, and machine learning to analyze and classify customer sentiment as positive, negative, or neutral [1].

Machine learning has made significant progress in classifying customer sentiment in online reviews. Unfortunately, the main focus of this study has concentrated on determining binary classification: positive and negative sentiment. The neutral sentiment is frequently ignored or removed. In an online review, a neutral class is derived from a 3-star rating for a customer who is satisfied enough with the quality of the product. For example: *"Would like to see better battery life. Decent otherwise."* The word "decent" lacks strong positive or negative sentiment about the product being reviewed. This review categorizes it as a neutral class with a mixture of both negative and positive feedback. Neutral class in customer sentiment analysis that allows comments data with factual, informative, or descriptive text rather than expressing a particular emotional tone.

Neutral sentiments should be handled or treated correctly to leverage the comments left and pinpoint areas of improvement. Many businesses pay close attention to reviews on both ends of the scale (4, 5 stars rating as positive and 1, 2 stars as negative) but must catch up on the valuable middle parts (3 stars as neutral). The number of 3 stars in online reviews might be small in volume but higher in value. According to Al-Rubaiee et al. [2], the neutral class can be helpful in sentiment analysis to maintain the accuracy of sentiment analysis models by reducing the risk of misclassifying text that does not express a clear sentiment.

From the business's viewpoint, neutral sentiment should be managed or addressed correctly to acquire a complete understanding of consumer feedback and find areas for improvement to make the product or services stand out. Businesses must guarantee that neutral sentiments are appropriately fulfilled to avoid consumer dissatisfaction, meet baseline requirements, and make well-informed decisions to encourage business growth.

The complexity of natural languages and the difficulty of quantifying human feelings make sentiment analysis a challenging task to categorize text data into different emotion classes automatically. Currently, research on sentiment classification is dominated by two basic approaches: machine learning and lexicon-based approaches. The machine learning approach requires a lot of labeled data training with manual process annotation, reviewing, and verification, which can slow system development and deployment.

Manual labeling using rating values were 1, 2 stars as negative sentiment, 3 stars as neutral sentiment, and 4-5 stars as positive sentiment is not possible to label the review sentences because customers often give a rating that does not match the review. For example, if the experience is valued with 5 stars (excellent), the sentence review has a negative connotation and vice versa. Some customer has their own biases when writing a review and these biases in rating may result in inconsistent reviews. Hu et al. [3] also claimed that customers are not entirely rational, and that affects self-selection rating biases.

The number of customer reviews for a popular product can be hundreds or thousands. Manual labeling commonly used in sentiment analysis is considered inefficient in terms of time and cost, especially if the data is extensive. Conversely, the lexicon-based approach can deal with language complexity and focuses on words and phrases as indicators of semantic orientation. However, this approach is low in accuracy but is computationally efficient, scalable, and provides consistent performance [4]. Combining lexicon-based and machine-learning techniques, known as the hybrid approach, can enhance performance accuracy in sentiment analysis results [5].

Multiple machine learning and lexicon-based approaches have been used to perform automatic classification for sentiment analysis. Unfortunately, the relative effectiveness of each approach is still unclear. This study employed a hybrid approach integrating the VADER lexicon with Multinomial Logistic Regression to examine the sentiment polarity of online customer reviews. The VADER methodology was selected as a lexicon-based approach that uses labels assigned in customer reviews for sentiment classification, utilizing a pre-defined vocabulary. Combined with Multinomial Logistic Regression as a supervised machine learning classifier, it uses a labeled sentiment from the VADER lexicon to train classifiers to generate more accurate sentiment predictions. According to Ramya and Rao [6], Multinomial Logistic Regression can make predictions in big data sets containing various diverse domains.

This research contributes to an extensive experiment with VADER and Multinomial Logistic Regression to determine how this combination affects the performance accuracy to detect neutral classes in sentiment classification. This process used the Amazon product review dataset to ensure the proposed approach is functional and executable in the customer sentiment domain. The performances, advantages, and limitations of VADER with Multinomial Logistic Regression in multiclass sentiment classification tasks were investigated and evaluated in this study. The TF-IDF method was selected as features vectorization to determine the important words to predict the sentiment in the Multinomial Logistic Regression algorithm.

Details about the relevant theories and related works are presented in Section II. Section III presents the research methodology for implementing a hybrid sentiment analysis approach. Experiment, results, and discussion are provided in section IV. Finally, Section V concludes with a conclusion and future works.

II. Related Works

A. Lexicon-based Approach with VADER

Sentiment classification uses automatic algorithms to predict the sentiment orientation of opinions included within a written document, such as a product review, blog post, or social media comment. Sentiment orientation can be characterized as positive, negative, or neutral, or it can be scored on a scale. The lexicon-based approach, also known as the dictionary-based approach, is employed in Natural Language Processing (NLP) for sentiment detection and text classification. This approach involves utilizing a predetermined lexicon or dictionary to identify specific words within the text [7]. A dictionary is prepared in advance, including entries for words or phrases linked to specific categories or sentiments. The lexicon can be established manually by domain experts or generated using automated approaches like machine learning. A score or label reflecting sentiment or category is given to each word or phrase in the lexicon. By aggregating the scores or labels of all the words or phrases in the text, an overall sentiment or category can be determined based on the number and intensity of positive and negative words encountered.

In this research, Valence Aware Dictionary and sEntiment Reasoner (VADER) is used in a lexicon-based approach. VADER is a lexicon-based algorithm developed by Hutto and Gilbert [8] in 2014 to solve the problem of analyzing language, symbols, and style of texts in sentiment analysis. It is widely used in various applications such as social media monitoring, customer feedback analysis, and brand reputation management. It utilizes a pre-built lexicon that contains words with sentiment scores and incorporates grammar and syntactical rules to handle negations, intensifiers, and modifiers [9].

VADER utilizes a pre-built lexicon that contains words or phrases with sentiment scores ranging from -1 (extremely negative) to +1 (extremely positive). The lexicon also includes words with neutral sentiment scores. The sentiment scores in VADER are based on human-annotated ratings and consider the intensity of sentiment associated with each word. The main output of VADER is a sentiment polarity score, which represents the overall sentiment expressed in each text. The score is a continuous value ranging from -1 to +1, where negative values indicate negative sentiment, positive values indicate positive sentiment and values close to zero indicate neutral sentiment [8].

Shabi [10] evaluates the performance of five lexicons used in sentiment analysis on Twitter data: VADER, SentiWordNet, SentiStrength, Liu and Hu opinion lexicon, and AFINN-111. By using the Stanford dataset, the results showed that the best performance in terms of accuracy was achieved by the VADER lexicon 72%, while the performance accuracy of the SentiStrength (67%), AFINN-111 (65%), Liu-Hu lexicon (65%), and SentiWordNet lexicon fell to the value 53%. With this comparison, the lexicon VADER has a good possibility for classifying short text pre-processing and can deal with multi-class, such as positive, negative, and neutral.

Furthermore, Heaton [11] conducts a comparative analysis of TextBlob and VADER in terms of sentiment analysis for positive, negative, and neutral sentiments expressed in social media regarding the NHS Covid-19 applications. The findings indicate that the VADER method outperforms recognized positive sentiments in sarcastic tweets with values of 0.8316, 0.7622, 0.6767, and 0.6958 for four tweets.

B. Machine Learning Approach with Multinomial Logistic Regression

Machine learning is an approach that creates models and designs algorithms to facilitate computational learning and decision-making processes. Machine learning has progressed beyond teaching computers to imitate the human brain to discovering statistical patterns in learning processes to deliver insights from datasets [12]. Machine learning can be classified into two categories: supervised and unsupervised learning. Supervised learning involves training a model using labelled data, where corresponding target labels or outcomes match the input data. These models are evaluated based on predictive capacity and variance measures. The objective is to acquire the knowledge necessary to construct a mapping function to accurately forecast the appropriate label for data sets known as testing data (unseen inputs). Multinomial Naïve Bayes, Support Vector Classification, Logistic Regression, Neural Networks, Random Forest, AdaBoost, Gradient Boosting, and Decision Tree are classifier models of supervised learning algorithms for multiclass classification.

In unsupervised learning algorithms, on the other hand, the model learns from unlabeled data without any predefined target labels. The objective is to discover patterns, structures, or relationships within the data by automatically developing classification labels by searching for similarities between data pieces to determine the category and create groups or clusters [13]. Clustering algorithms, such as K-means clustering, Latent Dirichlet Allocation (LDA), and Principal Component Analysis (PCA), are common algorithms of unsupervised learning. In this research, Multinomial Logistic Regression is selected as classifier algorithms to predict sentiment into positive, negative, and neutral classes in sentiment analysis.

Ramadhan et al. [14] applied Multinomial Logistic Regression in social media for Jakarta Governor Election with K-Fold cross-validation, achieving 74% accuracy with 90:10 training and testing data ratio. In his study, features were extracted and transformed into binary vectors using the TF-IDF method with the training dataset labeled manually. In addition, Purwandari et al. [15] compared Multinomial Logistic Regression and Multinomial Naïve Bayes for classifying weather in social media into five classes: cloudy, sunny, rainy, heavy rain, and light rain. The Multinomial Logistic Regression model has higher performance, with an accuracy rate of 83.3% and a precision rate of 90.3%. In comparison, the Multinomial Naïve Bayes model achieves an accuracy rate of 73.5% and a precision rate of 86.3%. Multinomial Logistic Regression has proven effective in classifying weather with good results.

C. *Hybrid Approach in Sentiment Analysis*

The hybrid approach is a methodology that integrates a lexicon-based approach and a machine-learning approach. Research investigations have recently focused on implementing a hybrid approach due to their ability to produce improved performance over either the lexicon-based model or the machine learning-based approach alone. Combining the advantages of lexicon-based and machine-learning methodologies is the primary goal of utilizing a hybrid method. The lexicon-based method is effective and can be used in a variety of contexts. This method does not require a significant amount of human interaction to label the training dataset and its ability to find the opinion words with the specific content orientation. On the other side, the machine learning approach effectively discovers subjectivity issues, noise resistance, and the ability to analyze numerous categories [16].

Most of the work in sentiment analysis performs supervised machine learning in binary classification with the best algorithms achieving an accuracy of less than 90%. A study by Pang et al. [17] demonstrated that the Support Vector Machine was 82.9% more accurate than the Naïve Bayes method in a movie review with binary classification. Vyas & Uma [18] observed that Support Vector Machine outperforms the Naïve Bayes and Decision Tree for binary sentiment classification in social media with an accuracy of 82.61%. Moreover, a study by Gupta et al. [19] discovered that the Random Forest algorithm outperforms Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor (KNN) with an accuracy rate of 78% for binary customer sentiment classification using Amazon, Yelp, and IMDB dataset.

There is a minimum study for a hybrid approach in multiclass sentiment classification, especially for VADER, a lexicon-based approach with a supervised machine learning classifier. Chaithra [20] analyzed the metadata of media-sharing sites (YouTube) with popular videos using a hybrid approach. The lexicon-based approach VADER was applied, and a Naïve Bayes classifier as machine learning was trained with 70% of the data. The classifier achieved an accuracy of 79.78% and an F1 Score of 83.72% on 30% testing data. By applying this approach, the accuracy value indicates that the text comment can be classified into positive or negative feedback rather than like/dislike on the YouTube site.

To add on, Mahmood et al. [21] employed a hybrid approach to classifying public opinion on social media in positive and negative sentiment, combining a lexicon-based with machine learning approaches such as Naïve Bayes and Support Vector Machines. The Support Vector Machine performs superior to the Naïve Bayes classifier, achieving an accuracy rate of 80% before combining with the lexicon-based approach, while the lexicon-based method alone reached 85%. The accuracy performance was increased after combining the lexicon-based and machine-learning approaches with a 90% accuracy rate.

According to the study by Rajeswari [4], observed multiclass classification using SentiWordNet with Logistic Regression for movie datasets achieved an accuracy of 89% compared to Naïve Bayes and K-NN classifiers. In addition, Mujahid et al. [22] applied a hybrid approach to classify tweet e-learning implementation using VADER, TextBlob, and SentiWordNet combined with Logistic Regression, SVM, K-NN, and Random Forest. Their study showed that VADER and Random Forest outperformed with an accuracy of 88%. Not only that, Sham and Mohamed [23] used a hybrid approach to classify sentiment into tri-class (positive, negative, neutral) using climate change tweets. The study found that hybrid approaches, including Logistic Regression and TextBlob using TF–IDF, outperformed on a combined dataset with an F1-score result of 75.3%. Lemmatization techniques are not recommended during the data preprocessing phase of lexicon approaches, as they can decrease the performance obtained. This study also found that

TF–IDF as the feature extraction technique outperformed Bag-of-Words (BoW) when used in Logistic Regression.

III. Proposed Methodology

This section provides an overview of the proposed methodology employed with a hybrid approach for multiclass sentiment analysis. The sequential execution of the proposed methodology is illustrated in Fig. 1.

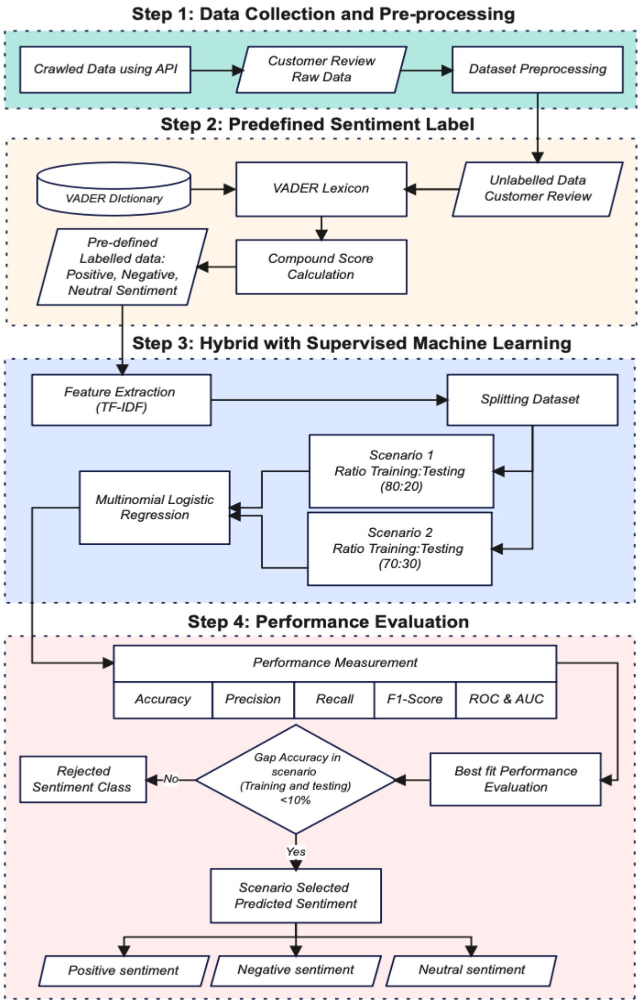

Step 1: Data Collection and Pre-processing
Crawled Data using API
Customer Review Raw Data
Dataset Preprocessing
Step 2: Predefined Sentiment Label
VADER Dictionary
VADER Lexicon
Unlabelled Data Customer Review
Pre-defined Labelled data: Positive, Negative, Neutral Sentiment
Compound Score Calculation
Step 3: Hybrid with Supervised Machine Learning
Feature Extraction (TF-IDF)
Splitting Dataset
Scenario 1 Ratio Training:Testing (80:20)
Multinomial Logistic Regression
Scenario 2 Ratio Training:Testing (70:30)
Step 4: Performance Evaluation
Performance Measurement
Accuracy
Precision
Recall
F1-Score
ROC & AUC
Rejected Sentiment Class
No
Gap Accuracy in scenario (Training and testing) <10%
Best fit Performance Evaluation
Yes
Scenario Selected Predicted Sentiment
Positive sentiment
Negative sentiment
Neutral sentiment

Fig. 1. Proposed methodology.

The first task is pre-processing, which includes transforming the case, deleting characters, and removing stop words to remove unneeded and repetitive information. Using the VADER lexicon, the second stage categorizes the customer review as positive, negative, or neutral. The third step is to use TF-IDF Vectorizer as a features extraction approach to execute Multinomial Logistic Regression for sentiment classification. The final step is to assess performance using accuracy, F1-Score, AUC, and ROC metrics.

A. Step 1: Dataset Collection and Pre-processing

1) Dataset: This study uses the Python programming language in a Jupyter Notebook to develop the suggested hybrid model. A web crawler was used to gather online customer reviews from Amazon.com, and mobile phone reviews were selected as a dataset with a more significant number of reviews. There are 3984 records of customer reviews with the following four attributes:

- Rating: consist of the customer assessment with range of 1 up to 5 stars to describe satisfaction level with the products. Rating on scale means: 1 - highly dissatisfied, 2 - dissatisfied, 3 - neutral, 4 - satisfied, and 5 - highly satisfied.
- Posting time: the date in customer post the review.
- Customer account: representing the customer identity in product review.
- Sentence of review: the text comments by customer after purchasing product to review the product performance.

All available information was crawled using a Python program. The Amazon website provides a review and rating score from 1-5 stars after customers buy products. Fig. 2 presents detailed rating information for the dataset.

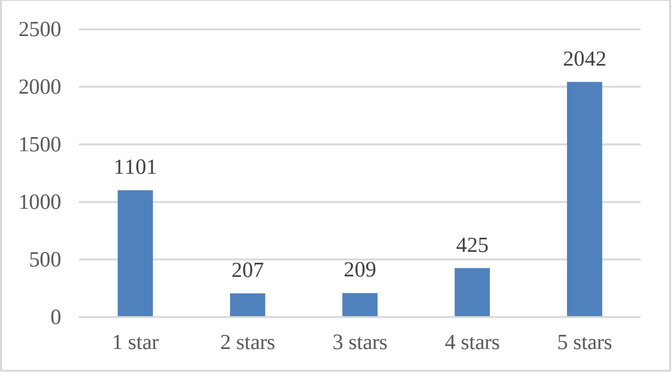


Fig. 2. Rating Information of mobile phone review

According to Fig. 2, the product has 1308 negative comments (33%) with 1-2 stars, 209 neutral sentiments (5%) with three stars, and 2467 with positive comments (62%) with 4-5 stars. This sentiment distribution indicates a positive product experience.

2) Pre-processing: Text pre-processing is important in sentiment analysis tasks due to the high dimensionality, poorly structured, and unstandardized text data tasks [24]. Pre-processing methods were performed to remove irrelevant and redundant data, which significantly impacted data quality. Pre-processing tasks include (1) case transformation, (2) tokenization, (3) stop word removal, and (4) stemming. Transform case is converted text to lowercase to ensure uniformity and prevent treating the same term differently due to case. The second procedure, tokenization, breaks textual data into smaller and meaningful components called tokens. Punctuation marks are removed during tokenization. The next step is to remove stop words. This method removes words from the text that don't add useful information, such as determiners, prepositions, coordinating conjunctions, and more. Finally, stemming is applied to the reduction of words to their roots.

B. Step 2: Predefined Sentiment Label

The customer review dataset is labeled as positive, negative, or neutral using VADER vocabulary sentiment. This labeling procedure makes training data for the supervised machine-learning model more efficient. VADER lexicon provides sentiment strength based on the polarity scores for each data review with an intensity value that falls between the range of -1 to 1. For example, the words *"perfect"* and *"great"* will have the same positive polarity, whereas in the VADER lexicon, "great" is more positive than "perfect" with the intensity value (valence score) for *great (0.79)* higher than the intensity value of *perfect (0.69).*

The sentiment intensity analyzer in the NLTK package, known as VADER, produces a sentiment score in a dictionary with four terms: neg, neu, pos, and compound. The terms "neg," "neu," and "pos" are used to represent negative, neutral, and positive meanings accordingly. The total of the numbers should approximate or equal to 1. The compound sentiment score is determined by summing the valence scores of every word in the lexicon, and it serves as an indicator of the overall sentiment intensity. The emotion score ranges from (-1), indicating the most extreme negative sentiment, and (+1) the most extreme positive attitude. The experiment utilized the compound score threshold value to ascertain the inherent sentiment of a given text, as described in Table I and applied in Python using Algorithm 1.

TABLE I. VADER LEXICON COMPOUND SCORE

Sentiment Polarity	Range
Positive	Compound score ≥ 0.05
Negative	*Compound Score* ≤ -0.05
Neutral	$-0.05 <$ compound score < 0.05

In order to implement the VADER algorithm in NLTK, it is necessary to download the VADER lexicon data and execute the commands using the Python script. The downloaded VADER lexicon was employed to apply the Sentiment Intensity Analyzer class from the NLTK for sentiment analysis. The Sentiment Intensity Analyzer class utilizes the polarity scores approach to provide a dictionary of sentiment scores. These scores include the compound score, a normalized composite value ranging from -1 to +1, and the positive, negative, and neutral scores. Based on specific requirements, the compound score threshold can be adjusted to categorize sentiment as positive, negative, or neutral.

Algorithm 1: The Sentiment VADER

```
Input: Pre-processed customer review data
Process:
    import nltk
    nltk.download('vader_lexicon')
    from nltk.sentiment import SentimentIntensityAnalyzer
    # Create an instance of the SentimentIntensityAnalyzer
    analyzer = SentimentIntensityAnalyzer()
    # Analyse sentiment of customer review sentence
    sentence = "I love everything about the phone!"
    sentiment_scores = analyzer.polarity_scores(sentence)
        if sentiment_scores['compound'] >= 0.05:
            sentiment = "positive"
        elseif sentiment_scores['compound'] <= -0.05:
            sentiment = "negative"
        else:
            sentiment = "neutral"
Output:
    # Print sentiment scores
    print sentiment_scores['compound']
```

C. Step 3: Hybrid with Supervised Machine Learning

The VADER-labeled data is paired with a supervised machine learning method employing Multinomial Logistic Regression to make predictions about the sentiment of customer reviews. Fig. 3 depicts the operational sequence of the Multinomial Logistic Regression model. The model is trained using a pre-defined VADER-labeled dataset. This dataset is divided into two different training and testing datasets scenario proportions.

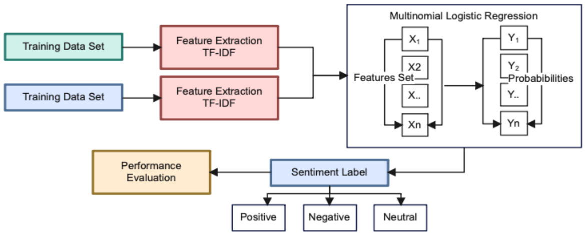


Fig. 3. Sentiment model with multinomial logistic regression.

1) Ratio proportion training and testing dataset: Training and testing data are split 80:20 and 70:30, respectively. In the first scenario, the 80% dataset is used for training and 20% for testing. In the second scenario, the 70% dataset is used for training and 30% is used for testing. This ratio was used when the dataset was large enough to provide sufficient training and testing instances. A larger training dataset can allow the model to learn more complex patterns, while smaller testing sets may result in less reliable performance estimates.

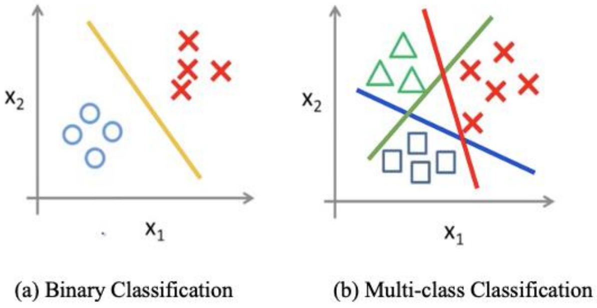
2) Feature extraction: This process aims to extract important and informative features from the cleaned dataset to reduce dimensionality and improve model performance and interpretability. The TF-IDF method was used in the feature extraction procedure to generate a document term matrix that denotes each term's word count and weightage [25].

$$\text{TF-IDF} = \text{tf x } 1 + \log\frac{N}{1+df(w)} \qquad (1)$$

The term frequency **(tf)** is calculated as the value of that term's occurrence in specific documents. The total number of documents in the corpus is denoted by **(N)**, while **df(w)** signifies the count of documents containing the term **w**.

3) Sentiment classification model: There are two types of classification issues based on the number of classes: binary classification and multi-class classification (more than two classes). In this study, a neutral sentiment class was handled using multiclass classification. Multiclass classification enables finer-grained data analysis by considering many categories, resulting in a deep insight and understanding of the data.

By default, Logistic Regression is a classification algorithm for binary classification. The positive (true) class is allocated a value of 1, and the negative (false) class is assigned a value of 0. The fit model predicts the likelihood of a class 1 [26]. Multinomial Logistic Regression, or extended logistic regression, predicts more than two classes. Fig. 4 depicts the distinction between binary and multiclass classification.



(a) Binary Classification (b) Multi-class Classification

Fig. 4. Binary and multi-class classification.

A commonly used method for extending binary Logistic Regression to handle multiclass classification problems involves dividing the multiclass problem into several binary classification problems and applying a standard Logistic Regression model to each individual problem. This technique is called *one-vs-rest* and *one-vs-one* wrapper models. In the *one-vs-rest* or *one-vs-all* approach, build a Logistic Regression to find the probability the observation belongs to each class. Fig. 5 describes the step of multiclass Logistic Regression.

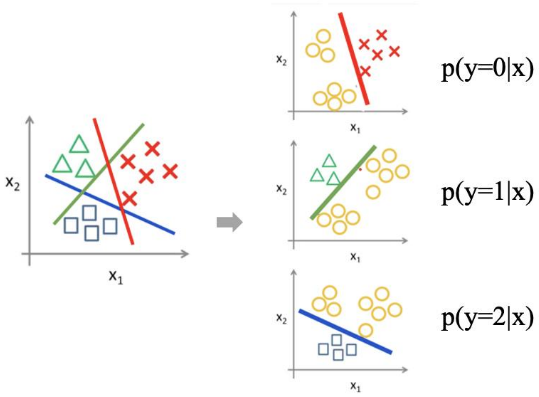


Fig. 5. Multiclass logistic regression classification.

The one-vs-rest technique involves training multiple binary Logistic Regression models, where each model represents one class against the rest of the classes [27]. The probabilities obtained from binary models combined to make multiclass classification predictions.

Denoted:

- N as the number of instances in the dataset
- K as the number of classes
- X as the input matrix of size N x D, where D is the number of features
- Y as the target variable matrix of size N x K, where each row represents the class labels for an instance using one-hot encoding

Train K binary logistic regression models, where each model i (i = 1 to K) predicts the probability of instance X belonging to class i against all other classes. The formula for the predicted probability of instance X belonging to class i is:

$$P(Y = i \mid X) = \sigma(W_iX + b_i) \qquad (2)$$

where:

- P(Y = i | X) is the predicted probability of instance X belonging to class i
- σ is the sigmoid function that maps the linear combination of the input features and model parameters to a probability between 0 and 1 (see Fig. 6).

$$\sigma\,(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+\exp(-z)} \qquad (3)$$

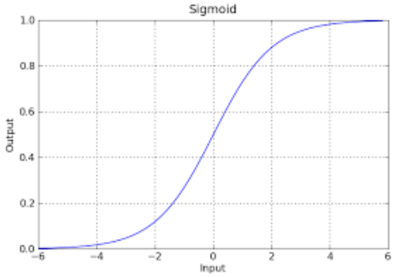


Fig. 6. Sigmoid function in logistic regression [27].

- W_i is the weight vector of size D x 1 for model i
- b_i is the bias term for model i

In order to generate predictions for a new instance X, the probabilities for each class must be computed utilizing the trained models, followed by the normalization probabilities with SoftMax formula [28]:

$$P(Y = i \mid X) = \frac{\exp(W_i X + b_i)}{\sum_{j=1}^{K} \exp(W_j X + b_j)} \quad 1 \leq i \leq K \qquad (4)$$

where:

- exp is the exponential function
- $\sum_{j=1}^{K} \exp(W_j X + b_j)$ is the sum of exponential terms for all classes j

Finally, the class label with the highest probability is assigned to the instance [29].

4) Model performance evaluation: The confusion matrix is a tabular representation that provides a detailed breakdown of a model's predictions. It displays the true positive, true negative, false positive, and false negative values for each class, allowing for an evaluation of the sentiment classification performance [30]. It helps to evaluate the correctness of classification approaches in multiclass classification problems. Table II presents the confusion matrix in multiclass classification with three sentiment categories: positive, negative, and neutral.

TABLE II. CONFUSION MATRIX FOR THREE SENTIMENT CLASSES

Actual	**Prediction**		
	Positive	***Negative***	***Neutral***
Positive	**TP**	**FNg1**	**FNt1**
	True Positive	False Negative 1	False Neutral 1
Negative	**FP1**	**TNg**	**FNt2**
	False Positive 1	True Negative	False Neutral 2
Neutral	**FP2**	**FNg2**	**TNt**
	False Positive 2	False Negative 2	True Neutral

where:

- True Positive (TP) refers to the number of times the classifier correctly predicts that the positive class is positive.
- The term True Negative (TN) refers to the number of the classifier that accurately predicts the negative class as negative.
- The term False Positive (FP) refers to the number of the classifier makes an incorrect prediction by classifying a negative class as positive.
- The number of the classifier incorrectly predicts the positive class as negative, referred to as False Negative(FN**).**

Performance evaluation metrics such as accuracy, precision, and recall can be formulated as follows by using a confusion matrix in Table II.

Accuracy is the percentage of correctly predicted instances across all classes divided by the total number of instances in the dataset.

$$Accuracy = \frac{TP+TNg+TNt}{TP+FP1+FP2+FNg1+TNg+FNg2+FNt1+FNt2+TNt}\ x\ 100\% \qquad (5)$$

Precision is the percentage of model positive predictions that are true. Precision in a multiclass classification problem can be calculated separately for each class, including positive, neutral, and negative.

$$Precision\ Positive\ = \frac{TP}{TP + FP1 + FP2}\ x\ 100\% \qquad (6)$$

$$Precision\ Negative\ = \frac{TNg}{TNg + FNg1 + FNg2}\ x\ 100\% \qquad (7)$$

$$Precision\ Neutral\ = \frac{TNt}{TNt + FNt1 + FNt2}\ x\ 100\% \qquad (8)$$

Recall called a true positive rate or sensitivity. Recall is also applied for each class in a multi-class classification problem that measures the percentage of true positive predictions out of all positive instances.

$$Recall\ Positive = \frac{TP}{TP + FNg1 + FNt1}\ x\ 100\% \qquad (9)$$

$$Recall\ Negative\ = \frac{TNg}{FP1 + TNg + FNt2}\ x\ 100\% \qquad (10)$$

$$Recall\ \ Neutral = \frac{TNt}{FP2 + FNg2\ + TNt}\ x\ 100\% \qquad (11)$$

F1 Score is a metrics that combines precision and recall value. This value provides a balanced assessment of the model's performance for positive, negative, and neutral classes. The formula to calculate the F1 Score is as follows:

$$F1\ Score\ Positive\ = \frac{2\ x\ Precision\ Positive\ x\ Recall\ Positive}{Precision\ Positive + Recall\ Positive} \qquad (12)$$

$$F1\ Score\ Negative\ = \frac{2\ x\ Precision\ Negative\ x\ Recall\ Negative}{Precision\ Negative + Recall\ Negative} \qquad (13)$$

$$F1\ Score\ Neutral\ = \frac{2\ x\ Precision\ Neutral\ x\ Recall\ Neutral}{Precision\ Neutral + Recall\ Neutral} \qquad (14)$$

ROC Curve and AUC is a graphical representation of the model's performance across different classification thresholds.

IV. EXPERIMENT AND RESULT

This section presents the result of the experiments and evaluates the performance of VADER and Multinomial Logistic Regression.

A. VADER Result

The polarity results of the dataset are determined using the VADER vocabulary as presented in Table III, and an inconsistency rating was identified in the dataset review compared with the VADER result. The result of the VADER classification sentiment in positive, negative, or neutral was described in Fig. 7 compared with the distribution frequency of sentiment based on the rating score in the raw dataset. The VADER lexicon can detect the neutral sentiment by assigning the sentiment score a neutral range for a text that does not strongly express positive or negative sentiment.

TABLE III. VADER LEXICON COMPOUND SCORE

Sentence	Rating	Compound Score	VADER Polarity	Consistency
I love everything about the phone! It's in great condition	5	0.819	Positive	True
Bought for nephew	5	0.000	Neutral	False
Too difficult to set up	1	-0.3612	Negative	True
Like the style but soon as I turn it on it gets really hot	2	0.3612	Positive	False
It is a good phone but not a great one	3	0.000	Neutral	True
Battery capacity was at 85% when I got it. Disappointing	3	-0.4939	Negative	False

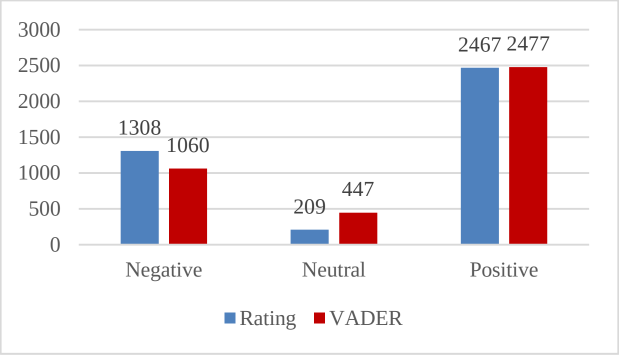


Fig. 7. Sentiment polarity distribution.

The accuracy rate for the VADER lexicon is 75.213% to label customer review data with sentiment polarity. VADER has been shown to perform effective in sentiment analysis tasks, particularly for customer reviews, where sentiments are often expressed in an informal and context-dependent manner.

TABLE IV. VADER LEXICON COMPOUND SCORE

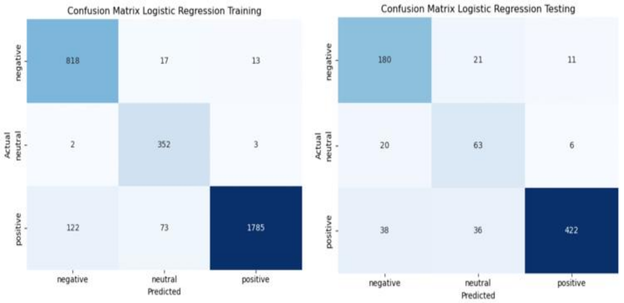
Bias Category from Rating to VADER sentiment	Total
Positive to Negative	**308**
Positive to Neutral	201
Negative to Positive	**263**
Negative to Neutral	206
Neutral to Positive	102
Neutral to Negative	**67**

According to Table IV, the investigation indicates misclassification between positive and negative and minimal misclassification between negative from neutral across all lexicons. VADER has a bias for inverting the polarity of ratings from positive to negative and vice versa. Since the differentiation between neutral with positive and neutral with negative are the significant factors to be clear in investigating the sentiment customer review, VADER lexicon is appropriate to adapted in task of labelling sentiment related to customer review.

B. *Sentiment Classification with Hybrid Approach: VADER with Multinomial Logistic Regression*

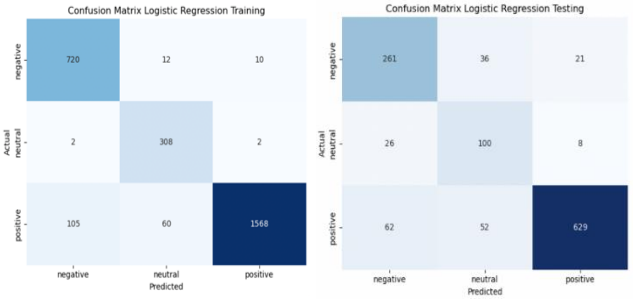
The metrics accuracy, precision, recall, F1-score, and Area Under Curve (AUC) are employed to assess the performance VADER with Multinomial Logistic Regression as a hybrid proposed model for multiclass classification. Confusion matrix visually shows the performance with two scenarios: splitting ratio 80:20 (see Fig. 8) and ratio 70:30 (see Fig. 9).

Regarding the result comparison in Table V, the first scenario with an 80:20 ratio is outperformed with an accuracy of 9,341% with a gap performance of <10% between training and testing data. Extending VADER with Multinomial Logistic Regression has increased the accuracy value by 17.565% from 75.213% with VADER to 92.778% with a hybrid approach. The second scenario has an overfitting condition with training data having a good performance, while its performance decreases significantly on the testing dataset (new data). Overfitting in machine learning should be avoided because it can negatively impact the model's performance and generalization ability on data that has never been seen before.



(a) Training data (b) Testing Data

Fig. 8. Confusion matrix with ratio 80:20.



(a) Training data (b) Testing Data

Fig. 9. Confusion matrix of with ratio 70:30.

TABLE V. GAP ANALYSIS FOR ACCURACY PERFORMANCE

Ratio	Dataset	True Predicted			Accuracy (%)	Gap Performance (%)
		Pos	***Neg***	***Neu***		
80:20	Training	1785	818	352	92.778	9,341
	Testing	422	63	180	83.437	
70:30	Training	1568	720	308	93.147	10.293
	Testing	629	261	100	82.845	

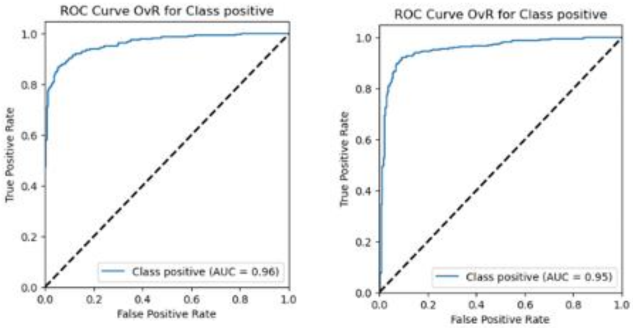
According to Table VI, the F1-Score for the positive, negative, and neutral classes are greater than 0.5 and close to 1. The positive classes have a higher score compared to the other classes. F1-Score combines precision and recall value works where the dataset is imbalanced, and according to this metric performance, a hybrid approach with a combination of VADER lexicon and Multinomial Logistic Regression accurately identifies positive, negative, and neutral classes with high recall and minimizes the frequency of false positives with high precision.

Furthermore, the AUC values in Table VI, which are close to 1, indicate that the hybrid approach is highly effective with excellent discrimination capabilities in distinguishing between positive, negative, and neutral classes. To better understand these data, the area under the ROC curve in Fig. 10 is represented by the area under the curve (AUC).

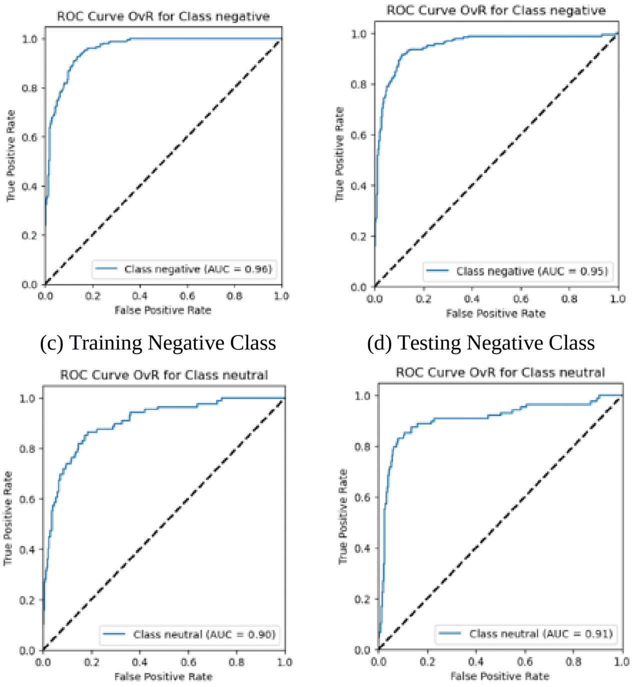
The ROC (Receiver Operating Characteristic) in Fig. 10 depicts a trade-off between the True Positive Rate and the False Positive Rate, which are shown on the "y-axis" and "x-axis," respectively. The line in the upper left corner of each ROC curve shows the cutoff value. The representation of ROC in Fig. 10 indicates that a hybrid approach with a combination of VADER lexicon and Multinomial Logistic Regression delivers a greater True Positive Rate (TPR) while preserving the low value in FPR.

TABLE VI. HYBRID APPROACH PERFORMANCE EVALUATION WITH SELECTED RATIO 80:20

Dataset		**Evaluation Metrics**			
		Precision	***Recall***	***F1-Score***	***AUC***
Training	Positive	0.991	0.902	0.944	0.96
	Negative	0.868	0.964	0.914	0.96
	Neutral	0.796	0.984	0.881	0.90
	Macro Avg	0.885	0.951	0.913	
	Weighted Avg	0.934	0.928	0.929	
Testing	Positive	0.961	0.851	0.903	0.95
	Negative	0.756	0.849	0.800	0.95
	Neutral	0.525	0.708	0.603	0.91
	Macro Avg	0.748	0.803	0.769	
	Weighted Avg	0.858	0.834	0.842	



(a) Training Positive Class (b) Testing Positive Class



(c) Training Negative Class (d) Testing Negative Class

(e) Training Neutral Class (f) Testing Neutral Class

Fig. 10. Hybrid approach ROC visualization (Ratio 80:20).

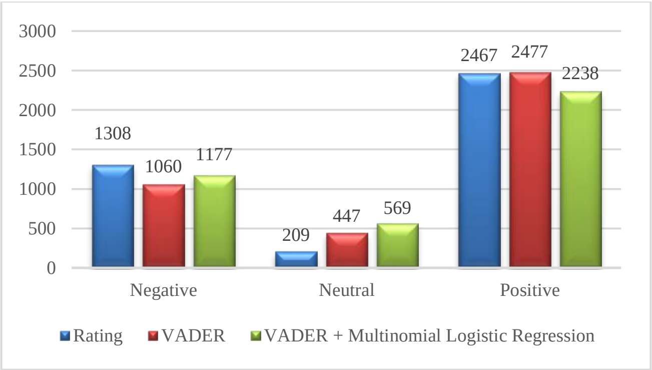


Fig. 11. Frequency distribution of sentiment polarity.

Fig. 11 presents the frequency distribution of sentiment polarity as determined by the rating score, VADER lexicon, and hybrid approach. The comparison results demonstrate that the hybrid approach effectively handles the neutral class when applied to multiclass sentiment classification. Extending VADER with Multinomial Logistic Regression can classify customer sentiment in online reviews as positive, negative, and neutral with good performance. The effectiveness of this method depends on the dataset accessibility, the complexity of sentiment nuances to be captured, and the specific criteria in the domain application.

V. CONCLUSION AND FUTURE WORKS

Online customer reviews can accurately predict customer sentiment regarding product experiences after purchase. With an accuracy percentage of 75.213%, VADER Lexicon classifies customer evaluations as positive, negative, or neutral, efficient in terms of time and costs for text labeling without a human annotator. The VADER model is interpretable and simple to use, but its functionality is limited by the lexicon. This method depends on words in the dictionary and may only work well with words in the lexicon. Additionally, the accuracy increased to 92.778% after combining with Multinomial Logistic Regression. The high level of accuracy indicates that the hybrid approach has an excellent performance in predicting the sentiment polarity in multiclass classification in customer reviews. The VADER model performs well in predicting the neutral class, whereas the Multinomial Logistic Regression model succeeds in an imbalanced dataset with high-dimensional features and overfitting challenges. Further investigation could involve conducting experiments to fine-tune the hybrid approach and comparing it to various algorithms in lexicon-based and machine-learning approaches.s

REFERENCES

[1] K. Jindal and R. Aron, “A Systematic Study of Sentiment Analysis for Social Media Data,” Materials Today Proceeding, Feb. 2021.

[2] H. AL-Rubaiee, R. Qiu, and D. Li, “The Importance of Neutral Class in Sentiment Analysis of Arabic Tweets,” International Journal of Computer Science and Information Technology, vol. 8, no. 2, pp. 17–31, Apr. 2016.

[3] N. Hu, P. A. Pavlou, and J. Zhang, “On self-selection biases in online product reviews,” MIS Quarterly, vol. 41, no. 2, pp. 449–471, 2017.

[4] A. M. Rajeswari, M. Mahalakshmi, R. Nithyashree, and G. Nalini, “Sentiment Analysis for Predicting Customer Reviews using a Hybrid Approach,” in Proceedings - Advanced Computing and Communication Technologies for High Performance Applications, Institute of Electrical and Electronics Engineers Inc., Jul. 2020, pp. 200–205.

[5] T. Sana, B. Ines, J. Salma, and Y. Ben Ayed, “A Hybrid Method for Arabic aspect-Based Sentiment Analysis,” International Journal Hybrid Intelligence System, vol. 16, no. 2, pp. 99–110, 2020.

[6] V. U. Ramya and K. T. Rao, “Sentiment Analysis of Movie Review using Machine Learning Techniques,” International Journal of Engineering & Technology, vol. 7, no. 2, pp. 676–681, 2018.

[7] S. Singh Hanswal, A. Pareek, and A. Sharma, “Twitter Sentiment Analysis using Rapid Miner Tool,” 2019.

[8] C. J. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” in Proceedings-International AAAI Conference on Weblogs and Social Media, Association for the Advancement of Artificial Intelligence, 2014, pp. 216–225.

[9] S. Panchal, (2020, March 7), “Sentiment Analysis with VADER- Label the Unlabelled Data,” Medium Website, Analytics Vidhya. https://medium.com/analytics-vidhya/sentiment-analysis-with-vader-label-the-unlabeled-data-8dd785225166

[10] M. A. Al-Shabi, “Evaluating The Performance of The Most Important Lexicons Used to Sentiment Analysis and Opinions Mining,” International Journal of Computer Science and Network Security, vol. 20, no. 1, pp. 51–57, 2020.

[11] D. Heaton, J. Clos, E. Nichele, and J. Fischer, “Critical reflections on three popular computational linguistic approaches to examine Twitter discourses,” PeerJ Computer Science, vol. 9, p. e1211, Jan. 2023.

[12] V. Nasteski, “An Overview of The Supervised Machine Learning Methods,” University St. Kliment Ohridski - Bitola, Dec. 2017.

[13] T. O. Ayodele, “Types of Machine Learning Algorithms,” InTech, Feb. 2010.

[14] W. P. Ramadhan, A. Novianty, and C. Setianingsih, “Sentiment Analysis Using Multinomial Logistic Regression,” in 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC), IEEE, Sep. 2017, pp. 46–49.

[15] K. Purwandari, T. W. Cenggoro, J. W. C. Sigalingging, and B. Pardamean, “Twitter-based Classification for Integrated Source Data of Weather Observations,” International Journal of Artificial Intelligence, vol. 12, no. 1, pp. 271–283, Mar. 2023.

[16] F. Hemmatian and M. K. Sohrabi, “A Survey on Classification Techniques for Opinion Mining and Sentiment Analysis,” Artificial Intelligence Review, vol. 52, no. 3, pp. 1495–1545, Oct. 2019.

[17] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment Classification using Machine Learning Techniques,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2002, pp. 79–86.

[18] V. Vyas and V. Uma, “An Extensive study of Sentiment Analysis tools and Binary Classification of tweets using Rapid Miner,” in Procedia Computer Science, Elsevier B.V., 2018, pp. 329–335.

[19] K. Gupta, N. Jiwani, and N. Afreen, “A Combined Approach of Sentimental Analysis Using Machine Learning Techniques,” Revue d’Intelligence Artificielle, vol. 37, no. 1, pp. 1–6, Feb. 2023.

[20] V. D. Chaithra, “Hybrid Approach: Naive Bayes and Sentiment VADER for Analyzing Sentiment of Mobile Unboxing Video Comments,” International Journal of Electrical and Computer Engineering, vol. 9, no. 5, pp. 4452–4459, 2019.

[21] A. T. Mahmood, S. S. Kamaruddin, R. K. Naser, and M. M. Nadzir, “A Combination of Lexicon and Machine Learning Approaches for Sentiment Analysis on Facebook,” Journal of System and Management Sciences, vol. 10, no. 3, pp. 140–150, 2020.

[22] M. Mujahid et al., “Sentiment Analysis and Topic Modelling on Tweets about Online Education During Covid-19,” Applied Sciences, vol. 11, no. 8348, pp. 2–25, Sep. 2021.

[23] N. M. Sham and A. Mohamed, “Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches,” Sustainability, vol. 14, no. 4723, pp. 1–28, Apr. 2022.

[24] C. Zhu and D. Gao, “Influence of data pre-processing,” Journal of Computing Science and Engineering, vol. 10, no. 2, 2016.

[25] M. R. Hasan, M. Maliha, and M. Arifuzzaman, “Sentiment Analysis with NLP on Twitter Data,” in 5th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, Institute of Electrical and Electronics Engineers Inc., Jul. 2019.

[26] J. Brownlee, (2020, September 1), “Multinomial Logistic Regression With Python,” Machine Learning Mastery. Accessed: Oct. 10, 2023. https://machinelearningmastery.com/multinomial-logistic-regression-with-python/

[27] “Introduction to Machine Learning Multi-class Classification,” PowerPoint slides. 2020. Carnegie Melon University.

[28] J. Daniel and J. H. Martin, “Logistic Regression,” in Speech and Language Processing, 2023, pp. 1–25.

[29] C. Wakamiya, “Classification with Logistic Regression,” PowerPoint slides. 2020. Berkeley SCET.

[30] A. E. S. Saputro, K. A. Notodiputro, and Indahwati, “Study of Sentiment of Governor’s Election Opinion in 2018,” International Journal of Scientific Research in Science, Engineering , and Technology, vol. 4, no. 11, pp. 231–238, Dec. 2018.

Artificial Intelligence-based Optimization Models for the Technical Workforce Allocation and Routing Problem Considering Productivity

Mariam Alzeraif, Ali Cheaitou
Department of Industrial Engineering and Engineering Management
University of Sharjah, Sharjah, UAE

Abstract—**Ensuring the reliability and availability of electric power networks is essential due to the increasing demands. An effective preventive maintenance strategy requires efficient resources allocation to perform the maintenance tasks, particularly the technical workforce. This paper introduces an innovative artificial intelligence-based approach to predict workforce productivity, aiming to optimize both the allocation of the technical workforce for maintenance tasks and their routing. In this study, two mathematical optimization models are introduced that utilize the output value of Artificial Neural Networks (ANN) for optimal resource allocation and routing. The first model focuses on team formation, considering the predicted productivity in order to ensure effective collaboration. While the second model focuses on the optimal assignment and routing of these teams to specific maintenance tasks. Validated with real-world data, the models show considerable promise in enhancing resource allocation, task assignment, and cost-efficiency in the electricity industry. Furthermore, sensitivity analysis has been conducted and managerial insights has been explored. The study also paves the way for future research, highlighting the potential for refining these models for more extensive applications.**

Keywords—Productivity; workforce; maintenance; optimization; allocation; routing

I. Introduction

The reliability and availability of the electric power network in the electricity industry is crucial for meeting the increasing demand [1]. Many of the blackouts were confirmed to be caused by imperfect maintenance due to human factors [2]. Maintenance can be divided into two categories: corrective and preventive. Corrective maintenance is performed after a breakdown. In contrast, preventive maintenance is performed at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure [3]. Consequently, maintenance planning, and more specifically preventive maintenance is extremely important in the electricity industry and plays a major role in reducing breakdowns and avoiding expensive blackouts.

For preventive maintenance tasks to be accomplished, resources and more specifically "technical workforce" must be allocated. Resource allocation refers to the decision-making process that determines the appropriate resource to perform each task or in other words, is described as "the best person for the task" [4]. In many cases, managers manually assign workforce to tasks based on intuition and experience [5]. In addition, workforce productivity is variable and can differ based on multiple factors and more importantly based on the maintenance task to be accomplished. Therefore, predicting the workforce productivity for each task based on relevant factors and considering the workforce productivity when assigning the maintenance tasks is important for better performance.

Furthermore, the geographically dispersed nature of electricity networks necessitates the movement of technical workforce between various locations to perform maintenance tasks. While there are no published records on the time technicians spend in transition and transfer between locations within the electricity industry, it is estimated based on the authors experience to range from 15 to 40% of working time. This can significantly increase costs and reduce operational efficiency in electricity companies. Therefore, upon allocating the tasks to the technical workforce based on predicted productivity, it is important to consider the routing of the allocated teams between the task sites at the same time as the time required to accomplish the tasks.

Although many research works stated that optimal resource allocation and routing will positively affect overall productivity, however, to the best of the authors' knowledge, no previous research clearly considered the predicted labor productivity as the main criterion in the human resources assignment and routing models as it will be shown in the following section.

The aim of this study is to allocate the technical workforce into teams and then assign the formed teams to preventive maintenance tasks and to plan their routes while considering the individual technical workforce productivity predicted using ANN. Therefore, to achieve this aim, the objective of this study is to develop two mathematical optimization models that provide optimal resource allocation and routing while considering the labor productivity produced by the ANN model as an input.

Fig. 1 presents the methodological approach adopted in this study. The first optimization model focuses on forming teams by pairing employees based on their productivity metrics predicted by the ANN model. This ensures that the teams are well-balanced and can operate at their maximum potential. Indeed, in electricity maintenance, tasks are required to be performed by teams of two employees at least for company's safety requirements. In addition, in the considered case study, teams include two employees each. The second optimization model focuses on the assignment of the formed teams to specific job-locations and therefore to optimize their routing. By considering various cost components, including wages, overtime cost, and transportation cost, the model aims to find the most cost-effective strategy for the entire operation. A numerical application based on a real-world data will be conducted to validate the developed models and the integration of the ANN output as an input for the optimization models. Finally, sensitivity analysis will be conducted to examine the impact of traffic, wages, and productivity on the optimal solution.

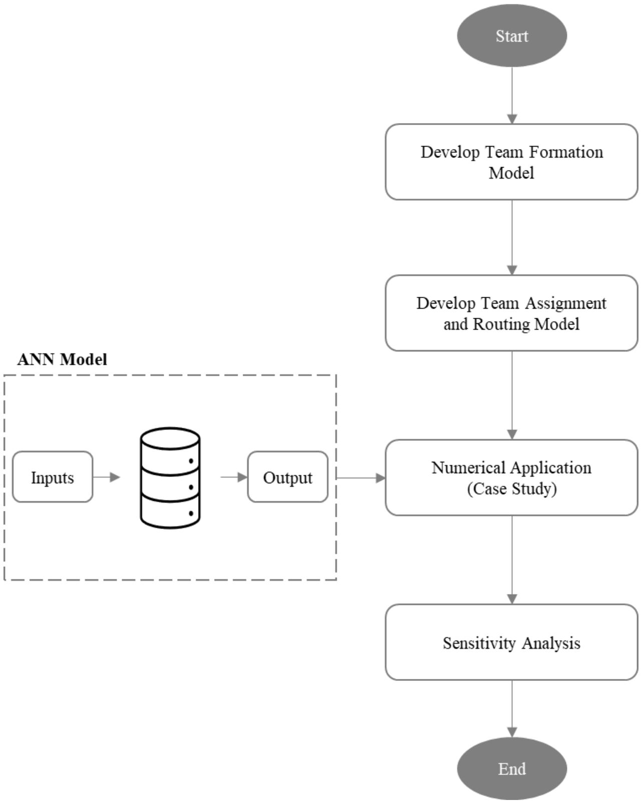

Start
Develop Team Formation Model
Develop Team Assignment and Routing Model
ANN Model
Inputs
Output
Numerical Application (Case Study)
Sensitivity Analysis
End

Fig. 1. The methodological approach.

The structure of this paper is organized as follows: Section II provides insight into resource assignment and routing problem related works, while Section III describes the mathematical formulation of the optimization models. Section IV focuses on the numerical application. Section V illustrates the sensitivity analysis. Finally, Section VI provides concluding remarks.

II. Literature Review

The assignment and routing of workforce is one of the important phases in the decision-making process, especially in the field of maintenance. For instance, a human resources assignment model was developed by [6] in the context of scheduling both planned and unplanned maintenance tasks. The human resources consist of three different specialties: plumber, electrician, and mechanic. The model takes into account the availability of human resources as well as support equipment, with the objective of maximizing the occupation rate of available human resources over the planning horizon. Similarly, the research work by [5] proposed a bi-objective model to assign licensed technicians to maintenance tasks across multiple work shifts. Each maintenance task must be assigned to only one technician, and only if the technician is licensed for that task. The model considers two objectives: the first is to minimize the cost for technicians to complete all the tasks, and the second is to minimize workload imbalances among technicians (ensuring workload fairness). The authors utilized a heuristic algorithm based on tabu search techniques to solve the proposed model.

In addition, a study proposed a dynamic maintenance task assignment model based on expert knowledge (experience), utilizing discrete stress–strength interference [7]. The authors employed the universal generating function method to calculate the value of experience. The objective is to determine which expert should be recommended for the corresponding maintenance task at various periods, based on the experts' values of experience, in order to maximize maintenance efficiency and reliability. A study presented the allocation and routing of technicians with different skills to perform maintenance tasks at offshore wind farms [8]. Similarly, another study proposed a manpower allocation problem in which teams of technicians with diverse skills are assigned a sequential order of tasks [9]. The model takes into account the time windows of tasks, the working hours of the staff, the skill requirements for tasks, and union regulations. A branch-and-price approach was used to solve the problem. Additionally, the same approach was employed to address the daily assignment of multi-skilled technicians into teams, tasked with maintenance along with routing and scheduling, with the objective of minimizing operational costs [10].

Another study examined the problem of maintenance planning for geographically distributed assets [11]. This research proposed a multi-objective model with the aim of minimizing total costs and maximizing the availability of the assets. The problem was addressed using a meta-heuristic solution method. The technician teaming and routing problem with service, cost, and fairness objectives was addressed in study [12]. They developed and solved mathematical optimization models for both an integrated and a sequential solution to the teaming and routing subproblems.

In the context of an electricity company, the study in [13] developed a model to assign maintenance task to a worker and to determine a schedule and route such that the downtimes of power lines and the travel effort of workers are minimized. The authors combine a Large Neighborhood Search meta-heuristic with mathematical programming techniques to solve the model. Similarly, research in [14] proposed a mixed integer programming model for multi-skilled technicians' assignment, along with the routing and scheduling problem. The aim is to form teams of technicians and assign them a sequence of planned and unplanned maintenance tasks to be performed within a given time window, depending on the type and urgency of the task. The objectives include completing higher priority tasks earlier and minimizing total operational costs.

A mixed-integer linear programming model was proposed by [15] to minimize costs associated with maintenance teams, spare parts, travel time, and noncompliance with service levels. The model was tested using various maintenance scenarios from a real maintenance provider in the UAE. The results demonstrated efficient time utilization, minimal routing schedules, and high service levels with a minimum number of teams. Recently, research in [16] explored the routing problem of preventive maintenance teams for elevator repair services, with the objective of minimizing penalties due to service earliness or lateness, assuming uniform travel times between nodes and team capability for any activity. The study developed a variable neighborhood search algorithm to solve the model.

Many of the human resource assignment problems applied in the maintenance field consider technicians' workload in the objective function. To the best of the authors knowledge, no works considered the specific problem of maintenance task assignment in the electricity industry based on labor productivity. Also, there have been no works related to this study that integrated ANN output with human resource assignment and routing problem as input.

III. Mathematical Models

A. Artificial Neural Networks Model

The ANN model has proven to be highly effective in predicting the productivity of the technical workforce, especially within the maintenance field of the electricity industry, as demonstrated in study [17]. In this study, we adopted the same configuration of the ANN model, which consisted of nine neurons in the input layer, one hidden layer with 15 neurons, and one neuron in the output layer. The activation function in the hidden layer is a sigmoid activation function (logsig), while the linear activation function (purelin) is in output layer. The model was trained using the backpropagation algorithm.

The input variables include type of equipment, employee skill level, employee health condition, level of safety measures, temperature, employee experience, level of supervisor competency, level of employee motivation or commitment, and humidity. The model aims to predict the productivity value per employee as an output parameter [17].

B. Team Formation Model

The daily planning of preventative maintenance tasks requires pairing technical workforce members into teams. Effective team formation is crucial for planning preventative maintenance to ensure tasks are executed efficiently and on schedule. This section presents the development of a team formation model, aiming to pair the members of the technical workforce into teams of two members each in a manner that minimizes disparity in productivity among them [12] [18]. The objective is not only to ensure that tasks are completed proficiently but also to prevent potential operational issues and inefficiencies that may arise from significant differences in team productivity.

The sets, parameters, decision variables, and mathematical model are described as follows:

1) Sets

$\mathcal{M}$: Set of employees indexed with $m, n = 1, \dots, M$.

$\mathcal{J}$: Set of jobs indexed with $j = 1, \dots, J$.

2) Parameters

P_{mj}: Productivity of employee m for job j; $m \in \mathcal{M}$; $j \in \mathcal{J}$.

AP_m : Average productivity of employee m; $m \in \mathcal{M}$. The average productivity is calculated using the following Equation:

$$AP_m = \frac{\sum_{j \in \mathcal{J}} P_{mj}}{J} \tag{1}$$

where:

$\sum_{j \in \mathcal{J}} P_{mj}$: is the summation of productivity values for employee m over all jobs in $\mathcal{J}$.

J: Total number of jobs.

3) Decision variables

X_{mn}: Binary variable where it is 1 if employees m and n are paired together in the same team and 0 otherwise; $m, n \in \mathcal{M}$.

dP_{mn}: A variable that represents the absolute difference in average productivity between employees m and n; $m, n \in \mathcal{M}$.

4) Mathematical model

$$\text{Minimize } \omega = \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{M}} dP_{mn} X_{mn} \tag{2}$$

Subject to:

$$\sum_{\substack{m \in \mathcal{M}, \\ m \neq n}} X_{mn} + \sum_{\substack{m \in \mathcal{M}, \\ m \neq n}} X_{nm} = 2 \qquad \forall n \in \mathcal{M} \tag{3}$$

$$X_{mn} = X_{nm} \qquad \forall m, n \in \mathcal{M} \tag{4}$$

$$dP_{mn} \geq (\ AP_m - \ AP_n\ ) X_{mn} \qquad \forall m, n \in \mathcal{M} \tag{5}$$

$$dP_{mn} \geq (\ AP_n - \ AP_m\ ) X_{mn} \qquad \forall m, n \in \mathcal{M} \tag{6}$$

$$X_{mn} \in \{0, 1\} \qquad \forall m, n \in \mathcal{M} \tag{7}$$

The objective function in Eq. (2) aims to minimize the sum of the absolute differences in average productivity between the members of all the formed teams. Constraint in Eq. (3) ensures that each employee is paired up with exactly one other employee. Constraint in Eq. (4) ensures the pairing is symmetrical; when an employee m is paired with an employee$n$, then n is paired with m. Constraint in Eq. (5) captures the non-negative difference in average productivity when employee m has a greater average productivity than employee n, and constraint in Eq. (6) captures the non-negative difference in average productivity when employee n has a greater average productivity than employee m. Together, Constraints in Eq. (5) and Eq. (6) ensure that dP_{mn} represents the absolute difference in average productivity between the two employees m and n. Constraint in Eq. (7) specifies ensures that the decision variables X_{mn} are binary.

C. Workforce Assignment and Routing Model

In the previous section, the team formation model was introduced. Now, the next stage is to assign teams to tasks and plan their routes in a way that minimizes the total cost [12]. Each team has a predicted productivity percentage per task, which affects the standard time required for the task. For example, if the team's predicted productivity is 80 percent, and the standard time for the task is 120 minutes, then the actual time will be 150 minutes, which is obtained by dividing the standard time by the productivity. Not all teams need to be assigned, but if a team is assigned, they are required to start their work from a designated depot and must return to it before the end of the regular working time. If a team returns after this designated time, overtime costs will be incurred.

The sets, parameters, decision variables, and mathematical model are detailed as follows:

1) Sets

$\mathcal{T}$: Set of teams indexed with $t = 1, \ldots, T$.

$\mathcal{J}$: Set of job-locations indexed with $i, j = 1, \ldots, J$.

$\mathcal{J}_0$: Set of job-locations to which the depot (where all teams start and end their routes) is added indexed with $i, j = 0, \ldots, J$.

It is worth noting each job-location is a location (i.e. a node) in the network at which only one job is to be performed. This is usual in electricity preventive maintenance where each location has a job performed by one team. Moreover, two different job-locations may correspond to the same type of maintenance tasks located in two different locations.

2) Parameters

dH_j : Standard time (minutes) required to complete the maintenance work of job-location j regardless of productivity of the team that will be assigned; $j \in \mathcal{J}$. It corresponds to the earned hours.

P_{tj} : Productivity of team t in performing job j; $t \in \mathcal{T}$; $j \in \mathcal{J}$. This value is derived from the ANN model mentioned in Section III.A [17]. It corresponds to the minimum productivity value of all the team members for job j.

F_t^v : Variable cost per hour of team t; $t \in \mathcal{T}$

F_{ij}^r : Transportation cost per team from job-location i to job-location j; $i, j \in \mathcal{J}_0$.

F_t^o : Overtime cost per hour for team t; $t \in \mathcal{T}$

T_{ij}^r : Time (Minutes) required to move from job-location i to job-location j; $i, j \in \mathcal{J}_0$.

$\mathcal{H}$: Number of regular working hours per day.

α : Maximum number of overtime hours per day for each team.

$\mathcal{S}$: Very large positive number.

3) Decision variables

Z_{tij}: Binary variable that is equal to 1 if team t performs the work in job-location (i.e., job) i directly before job j and 0 otherwise; $t \in \mathcal{T}$; $i, j \in \mathcal{J}$.

WT_t: Total number of regular working hours spent by team t per day; $t \in \mathcal{T}$.

O_t: Total number of overtime hours spent by team t per day; $t \in \mathcal{T}$.

f_j: Completion time of job-location j; $j \in \mathcal{J}$.

4) Mathematical model

Minimize $\mathcal{X}$

$$= \sum_{t \in \mathcal{T}} F_t^v \, WT_t + \sum_{t \in \mathcal{T}} F_t^o \, O_t + \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{J}_0} \sum_{j \in \mathcal{J}_0} F_{ij}^r \, \mathrm{Z}_{tij} \qquad (8)$$

Subject to:

$$\sum_{j \in \mathcal{J}} Z_{t0j} \leq 1 \qquad \forall t \in \mathcal{T} \qquad (9)$$

$$\sum_{i \in \mathcal{J}} Z_{ti0} \leq 1 \qquad \forall t \in T \qquad (10)$$

$$\sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{J}_0} Z_{tij} = 1 \qquad \forall j \in \mathcal{J} \qquad (11)$$

$$\sum_{i \in \mathcal{J}_0} Z_{tij} = \sum_{i \in \mathcal{J}_0} Z_{tji} \qquad \forall t \in \mathcal{T}, j \in \mathcal{J}_0 \qquad (12)$$

$$\sum_{i \in \mathcal{J}_0} \sum_{j \in \mathcal{J}_0} (T_{ij}^r + \frac{dH_j}{P_{tj}}) \times Z_{tij} = WT_t + O_t \qquad \forall t \in \mathcal{T} \qquad (13)$$

$$f_i + T_{ij}^r + \frac{dH_j}{P_{tj}} \leq f_j + \mathcal{S} \times \left(1 - Z_{tij}\right) \qquad \forall t \in \mathcal{T}, \; i \in \mathcal{J}_0, j \in \mathcal{J} \qquad (14)$$

$$f_j \geq 0 \qquad \forall j \in \mathcal{J}_0 \qquad (15)$$

$$WT_t \leq \mathcal{H} \qquad \forall t \in \mathcal{T} \qquad (16)$$

$$O_t \leq \alpha \qquad \forall t \in \mathcal{T} \qquad (17)$$

$$Z_{tij} \in \{0, 1\} \qquad \forall t \in \mathcal{T} \qquad (18)$$

The objective function in Eq. (8) minimizes the total cost in order to find the most cost-effective strategy for teams' assignment and routing, taking into consideration three pivotal cost components. The first component is associated with the cost of regular working hours of each team, capturing the standard wage or payment the team t would receive during regular working hours. The second component is associated with the cost of overtime hours of each team t. This factor represents the additional expense that might be incurred when teams work beyond standard working hours. The last component is the transportation cost for every move from location i to location j. In other words, the objective function ensures that tasks are allocated to teams and routed in the most economically efficient manner, balancing work hours, overtime, and travel expenses.

Constraints in Eq. (9) and Eq. (10) ensure that each team should start and end their route at the depot at most once. Constraint in Eq. (11) states that each job-location is visited once by one of the teams. Constraint in Eq. (12) guarantees that each team visiting a node j also leaves this node. Constraint in Eq. (13) specifies the value of the working hours and overtime hours of each team, which is composed of the traveled times and job time considering team productivity. Constraint in Eq. (14) defines the time at which job j is completed by team t. It requires that the completion time of the proceeding job i plus the travel time from i to j and the processing time of job j is a lower bound for the completion time of job j. In this constraint $\mathcal{S}$ denotes a sufficiently large positive value. Constraint in Eq. (15) ensures non-negative completion time. Constraint in Eq. (16) ensures that a team's regular working hours in a day do not exceed the maximum limit, while Constraint in Eq. (17) ensures that the overtime hours a team can work in a day should not exceed the maximum limit set by the organization's policy. Constraint in Eq. (18) specifies the domain of the decision variable.

IV. Numerical Study

In this section, the team formation and team assignment and routing models will be applied to the specific real-world context of preventive maintenance planning within a large electricity company that is responsible for generating, transmitting, and distributing power. Subsequent subsections will illustrate the details of the data collected. The numerical application serves as a demonstration of how these models can be effectively utilized to optimize preventive maintenance planning in the electricity industry and what kind of results can be expected.

A. Team Formation

1) Inputs data: In this stage, the focus is on forming teams based on the predicted productivity of each employee for each job. Initially, details about the maintenance jobs required to be performed are obtained. These details represent the maintenance jobs for a day within one zone. To ensure a robust validation of the model, the selected day had the highest number of jobs in that zone during a year. Subsequently, the ANN model developed in [17] and discussed above is employed to predict productivity of each employee for each maintenance job. The ANN model was executed after gathering and normalizing the necessary inputs data. Table I presents the predicted productivity values (P_{mj}).

2) Results: The developed mathematical model for team formation has been solved using IBM ILOG CPLEX optimization software, with a computation time of 1.05 seconds, using Intel Core i7 at 3.70 GHz computer with 16 GB RAM. the optimal team formation has been determined with an optimal objective value of $\omega^* = 0.277$. Table II displays the team formation results (X_{mn}).

TABLE I. Predicted Productivity Value Per Employee Per Job

j m	1	2	3	4	5	6	7	8	9	10	11
1	1.361	1.057	1.057	1.057	1.361	1.361	1.262	1.361	1.361	1.226	1.361
2	1.236	1.071	1.071	1.071	1.236	1.236	1.114	1.236	1.236	1.161	1.236
3	1.019	0.794	0.794	0.794	1.019	1.019	0.960	1.019	1.019	1.033	1.019
4	1.056	1.128	1.128	1.128	1.056	1.056	1.018	1.056	1.056	0.998	1.056
5	1.085	1.207	1.207	1.207	1.085	1.085	1.032	1.085	1.085	1.065	1.085
6	1.080	0.808	0.808	0.808	1.080	1.080	1.009	1.080	1.080	1.058	1.080
7	0.875	0.605	0.605	0.605	0.875	0.875	0.969	0.875	0.875	0.803	0.875
8	1.120	0.833	0.833	0.833	1.120	1.120	1.042	1.120	1.120	1.078	1.120
9	1.056	1.018	1.018	1.018	1.056	1.056	1.005	1.056	1.056	0.915	1.056
10	1.190	1.106	1.106	1.106	1.190	1.190	1.147	1.190	1.190	1.128	1.190
11	1.177	1.082	1.082	1.082	1.177	1.177	1.191	1.177	1.177	1.080	1.177
12	1.162	1.502	1.502	1.502	1.162	1.162	1.077	1.162	1.162	1.293	1.162
13	1.120	0.833	0.833	0.833	1.120	1.120	1.042	1.120	1.120	1.078	1.120
14	1.063	1.152	1.152	1.152	1.063	1.063	1.016	1.063	1.063	1.026	1.063
15	1.008	0.797	0.797	0.797	1.008	1.008	0.951	1.008	1.008	1.030	1.008
16	1.041	0.915	0.915	0.915	1.041	1.041	1.057	1.041	1.041	0.966	1.041

TABLE II. TEAM FORMATION RESULTS

No.	Employee (m)	Employee (n)	Absolute difference in average productivity between employees (dP_{mn})
1	1	12	0.002090909090908
2	2	10	0.015545454545455
3	3	6	0.043818181818182
4	4	14	0.012727272727273
5	5	11	0.031909090909091
6	7	15	0.143909090909091
7	8	13	0.000000000000000
8	9	16	0.026909090909091

B. Workforce Assignment and Routing

1) Inputs data: The data acquired regarding the maintenance jobs to be performed are utilized, along with additional information and details. Using the team formation results, team productivity values have been determined based on the lowest predicted productivity among its members. Table III displays the predicted productivity values of each team for each job (P_{tj}). The daily duration of regular working time for each team is set at eight hours per day ($\mathcal{H}$), while the maximum allowable overtime per team is limited to two hours per day (α). Team regular cost (F_t^v) and team overtime cost (F_t^o) in AED per hour are illustrated in Table IV. Moreover, Table V presents the standard time required to complete the maintenance jobs in minutes (dH_j). This standard time was established by the maintenance department of the selected company, assuming that all teams will perform at the same productivity level. Another critical dataset required for team routing is the transportation time matrix and transportation cost matrix. These matrices provide a systematic representation of time and cost associated with moving between different job-locations. Table VI displays (T_{ij}^r) the transportation times between job-locations that teams must traverse to perform maintenance tasks, while Table VII presents (F_{ij}^r) the transportation costs between these locations. The transportation cost is derived from the actual distances between job-locations obtained from [19]. For each kilometer of distance, the cost is calculated by considering both the average fuel expense and other vehicle-related expenses. It is worth noting that transportation time and cost between locations are not symmetric. For example, moving from job-location (2) to job-location (3) takes 19 minutes, but moving from job-location (3) to job-location (2) takes 29 minutes. Finally, the very large positive number ($\mathcal{S}$) has been set to 10,000.

TABLE III. PREDICTED PRODUCTIVITY VALUE PER TEAM PER JOB (P_{tj})

j t	1	2	3	4	5	6	7	8	9	10	11
1	1.162	1.057	1.057	1.057	1.162	1.162	1.077	1.162	1.162	1.226	1.162
2	1.190	1.071	1.071	1.071	1.190	1.190	1.114	1.190	1.190	1.128	1.190
3	1.019	0.794	0.794	0.794	1.019	1.019	0.960	1.019	1.019	1.033	1.019
4	1.056	1.128	1.128	1.128	1.056	1.056	1.016	1.056	1.056	0.998	1.056
5	1.085	1.082	1.082	1.082	1.085	1.085	1.032	1.085	1.085	1.065	1.085
6	0.875	0.605	0.605	0.605	0.875	0.875	0.951	0.875	0.875	0.803	0.875
7	1.120	0.833	0.833	0.833	1.120	1.120	1.042	1.120	1.120	1.078	1.120
8	1.041	0.915	0.915	0.915	1.041	1.041	1.005	1.041	1.041	0.915	1.041

TABLE IV. REGULAR COST (F_t^v) AND OVERTIME COST (F_t^o) PER TEAM

Team Number	Regular Cost [AED/Hour]	Overtime Cost [AED/Hour]
1	425.00	531.25
2	350.00	437.50
3	460.00	575.00
4	310.00	387.50
5	350.00	437.50
6	425.00	531.25
7	425.00	531.25
8	385.00	481.25

TABLE V. THE STANDARD TIME FOR EACH JOB-LOCATION (dH_j)

Job-Location	**Standard Time [Minutes]**
1	160
2	90
3	90
4	90
5	130
6	160
7	190
8	130
9	150
10	120
11	120

TABLE VI. TRANSPORTATION TIME MATRIX [MINUTES]

j *i*	**Depot**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Depot	0.0	33.0	20.0	34.0	35.0	38.0	37.0	39.0	39.0	40.0	33.0	34.0
1	35.0	0.0	28.0	20.0	12.0	22.0	17.0	17.0	22.0	21.0	17.0	25.0
2	17.0	25.0	0.0	19.0	23.0	23.0	25.0	25.0	27.0	26.0	23.0	20.0
3	37.00	19.00	29.0	0.0	12.0	11.0	12.0	16.0	11.0	13.0	21.0	13.0
4	35.00	15.00	28.0	12.0	0.0	15.0	10.0	10.0	16.0	14.0	14.0	20.0
5	36.0	26.0	31.0	12.0	17.0	0.0	13.0	16.0	6.0	14.0	20.0	17.0
6	38.0	17.0	25.0	12.0	11.0	13.0	0.0	7.0	14.0	10.0	15.0	23.0
7	36.0	19.0	28.0	14.0	12.0	15.0	6.0	0.0	15.0	13.0	17.0	23.0
8	41.0	25.0	31.0	12.0	18.0	3.0	15.0	17.0	0.0	12.0	22.0	20.0
9	39.0	23.0	33.0	12.0	13.0	15.0	9.0	13.0	16.0	0.0	17.0	22.0
10	36.0	11.0	25.0	19.0	14.0	21.0	15.0	18.0	21.0	20.0	0.0	25.0
11	26.0	19.0	18.0	13.0	16.0	18.0	16.0	18.0	18.0	18.0	17.0	0.0

TABLE VII. TRANSPORTATION COST MATRIX [AED]

j *i*	**Depot**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Depot	0.0	22.5	10.0	20.0	20.0	21.0	22.5	22.0	21.5	22.5	21.0	19.0
1	29.5	0.0	14.5	6.0	3.3	10.0	5.5	5.0	10.5	7.5	5.5	10.0
2	8.5	13.0	0.0	11.5	12.0	12.5	14.0	13.5	13.0	14.0	15.0	10.5
3	19.5	10.0	12.5	0.0	2.6	3.7	4.6	4.6	4.0	4.2	8.0	3.2
4	25.5	3.1	12.5	2.8	0.0	8.0	3.4	3.0	8.5	5.0	6.0	4.6
5	20.0	14.0	13.0	4.1	5.5	0.0	5.0	5.5	0.6	5.5	11.5	5.5
6	21.5	5.0	14.5	4.1	3.2	5.0	0.0	1.5	5.5	2.3	8.5	7.0
7	21.0	5.0	14.0	4.7	3.0	5.5	1.3	0.0	6.0	2.9	8.5	6.0
8	21.0	8.0	14.0	3.3	4.9	1.2	4.3	5.0	0.0	2.4	11.0	6.0
9	25.0	12.0	18.5	3.7	3.7	5.5	3.4	4.1	5.5	0.0	10.0	8.0
10	22.5	4.3	17.0	10.5	6.0	11.5	8.0	8.0	12.0	9.0	0.0	10.5
11	16.5	6.5	9.5	3.3	5.0	8.0	7.5	7.0	8.0	7.5	8.5	0.0

TABLE VIII. TEAM ASSIGNMENT AND ROUTING RESULTS

Team Number	**The Sequence of Performing Job-Location**	**Working Time [Minutes]**	**Overtime [Minutes]**
Team 2	Depot – 6 – 9 – 8 – 5 – Depot	480	100.99
Team 4	Depot – 2 – 3 – 4 – 7 – Depot	480	43.37
Team 5	Depot – 10 – 1 – 11 – Depot	465.74	0

2) Results: Using the developed mathematical model, the optimal assignment and routing have been obtained with a total optimal cost of 9,164.56 AED. Table VIII shows the sequence of job-locations to be visited and performed, total working time, and total overtime by each of the assigned teams. Note that five teams, specifically, teams number 1, 3, 6, 7, and 8 will not be assigned to any preventive maintenance task and will be supporting the emergency function. The model has been solved using IBM ILOG CPLEX optimization software, with a computation time of 14608.77 seconds, using Intel Core i7 at 3.70 GHz computer with 16 GB RAM.

V. Sensitivity Analysis

A. *Impact of the Trasportation Time (Traffic) on the Optimal Solution*

To study the impact of the transportation time (traffic) on the optimal solution, we increased the transportation time to consider different traffic scenarios, while maintaining the other parameters equal to their values considered in the original data set. The value of T_{ij}^{r} has been changed in each trial for selected routes based on the actual traffic condition corresponding to the expected time at which the team will travel from job-location i to j. The model has been solved for every change in the travelling time and the optimal solution has been obtained and noted. Fig. 2 identifies the effect of changing the transportation time (traffic) on the objective value (total cost). The results of each trial are detailed in Table IX. The sensitivity analysis reveals that the model is notably influenced by variations in transportation time, impacting not only the cost but also the optimal routing sequences. The working and overtime hours also showcased sensitivity to changes in transportation time, albeit with varied impact across different teams and trials.

TABLE IX. Detailed Results of the Effect of Traffic Time on Optimal Solution

Trial	Traffic Time [Minutes]	Total Cost [AED]	CPU Time [Seconds]	Team Number	Optimal Sequence of Job-Locations	Working Time [Minutes]	Overtime [Minutes]
1	18	9195.14	18360.31	Team 2	Depot – 5 – 8 – 9 – 6 – Depot	480	101.99
				Team 4	Depot – 3 – 4 – 7 – 2 – Depot	480	47.37
				Team 5	Depot – 10 – 1 – 11 – Depot	465.74	0
2	38	9216.30	26215.47	Team 2	Depot – 5 – 8 – 9 – 6 – Depot	480	101.99
				Team 4	Depot – 3 – 4 – 7 – 2 – Depot	480	47.37
				Team 5	Depot – 11 – 10 – 1 – Depot	467.74	0
3	90	9222.17	15859.78	Team 2	Depot – 8 – 5 – 9 – 6 – Depot	480	101.99
				Team 4	Depot – 4 – 7 – 3 – 11 – Depot	480	78.22
				Team 5	Depot – 1 – 10 – 2 – Depot	435.32	0
4	124	9320.80	24970.38	Team 2	Depot – 9 – 6 – 5 – 8 – Depot	480	107.99
				Team 4	Depot – 7 – 3 – 11 – 2 – Depot	480	81.22
				Team 5	Depot – 10 – 1 – 4 – Depot	441.32	0
5	202	9453.31	31804.34	Team 2	Depot – 1 – 8 – 5 – 11 – Depot	480	95.78
				Team 4	Depot – 2 – 3 – 4 – 7 – Depot	480	54.37
				Team 5	Depot – 10 – 6 – 9 – Depot	480	22.39

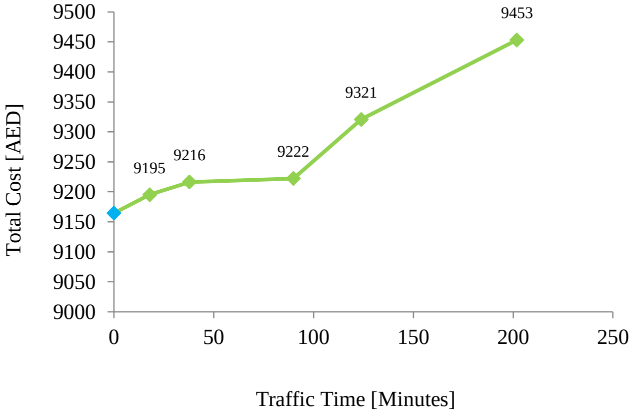


Fig. 2. The effect of traffic time on total cost.

B. *Managerial Insights*

The sensitivity analysis conducted underscores the substantial impact that transportation time can have on total costs. Notably, the correlation between increased transportation time and higher costs necessitates a proactive approach to traffic management by decision-makers. Managers should not only be vigilant about traffic conditions but also develop comprehensive strategies to mitigate their impact, as follows:

1) Avoiding peak traffic hours in the morning and afternoon. For example, teams could start at 6:00 AM instead of 7:30 AM.

2) Utilizing real-time traffic applications such as Google Maps can assist in identifying the most efficient paths and anticipating traffic bottlenecks, allowing for preemptive route adjustments.

3) Implementing flexible scheduling systems that adapt in real-time to traffic conditions can minimize unproductive hours spent in transit.

4) Considering the introduction of a second depot; the location of this depot should be informed by a thorough analysis of historical data on traffic patterns, job locations, and team movements to determine the most effective placement.

VI. Conclusion

In this study, team formation and team assignment and routing models that consider teams productivity based on ANN have been presented. The team formation model aimed to minimize the disparity in productivity among the members of each team. This is crucial in the context of maintenance as disparity can lead to inefficiencies, missed schedules, and potential operational disruptions. The objective of the team assignment and routing model was to minimize costs associated with regular working time, overtime, and transportation. Ensuring an economical approach to task assignment and routing is fundamental to operational efficiency, cost savings, and meeting maintenance schedules.

The real-world numerical application fortifies the practical applicability of the proposed models. In the context of an electricity company, the models showcased how one can optimize preventive maintenance planning, leading to improved resource utilization and cost-effectiveness. The rapid computation time, even for real-life sized scenarios, also underscores the feasibility of incorporating such models in everyday operational decisions. Furthermore, sensitivity analysis has been conducted to examine the effects of transportation time (traffic) on the optimal solution. It was noticed that increases in transportation time significantly raised total costs. The analysis revealed that longer travel times, often due to traffic congestion, directly correlate with increased expenses, highlighting the need for efficient route planning and traffic management strategies.

Future research may consider implementing many possible changes to the developed model. For Instance, includes outages time window. Additionally, future research should also aim to address the limitations of this study, such as the potential to use heuristics approaches in order to find the optimal solution for all the zones instead of only one zone at the same time.

Acknowledgment

The authors express their sincere gratitude to the experts who provided invaluable opinions and contributions to this study.

References

[1] A. Froger, M. Gendreau, J. E. Mendoza, É. Pinson and L.-M. Rousseau, "Maintenance scheduling in the electricity industry: A literature review," *European Journal of Operational Research*, vol. 251, no. 3, pp. 695-706, 2016.

[2] Y. Bao, C. Guo, J. Zhang, J. Wu, S. Pang and Z. Zhang, "Impact analysis of human factors on power system operation reliability," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 1, pp. 27-39, 2018.

[3] N. M. Paz and W. Leigh, "Maintenance scheduling: issues, results and research needs," *International Journal of Operations & Production Management*, vol. 14, no. 8, pp. 47-68, 1994.

[4] S. Bouajaja and N. Dridi, "A survey on human resource allocation problem and its applications," *Operational Research International Journal*, vol. 17, pp. 339-369, 2017.

[5] G. Chen, W. He, L. C. Leung, T. Lan and Y. Han, "Assigning licenced technicians to maintenance tasks at aircraft maintenance base: a bi-objective approach and a Chinese airline application," *International Journal of Production Research*, vol. 55, no. 19, pp. 5550-5563, 2017.

[6] D. Aït-Kadi, J. -B. Menye and H. Kane, "Resources assignment model in maintenance activities scheduling," *International Journal of Production Research*, vol. 49, no. 22, pp. 6677-6689, 2011.

[7] L. Li, M. Liu, W. Shen and G. Cheng, "An expert knowledge-based dynamic maintenance task assignment model using discrete stress–strength interference theory," *Knowledge-Based Systems*, vol. 131, pp. 135-148, 2017.

[8] A. H. Schrotenboer, M. A. uit het Broek, B. Jargalsaikhan and K. J. Roodbergen, "Coordinating technician allocation and maintenance routing for offshore wind farms," *Computers and Operations Research*, vol. 98, pp. 185-197, 2018.

[9] A. Dohn, E. Kolind and J. Clausen, "The manpower allocation problem with time windows and job-teaming constraints: A branch-and-price approach," *Computers & Operations Research*, vol. 36, no. 4, pp. 1145-1157, 2009.

[10] E. Zamorano and R. Stolletz, "Branch-and-price approaches for the Multiperiod Technician Routing and Scheduling Problem," *European Journal of Operational Research*, vol. 257, no. 1, pp. 55-68, 2017.

[11] M. Rashidnejad, S. Ebrahimnejad and J. Safari, "A bi-objective model of preventive maintenance planning in distributed systems considering vehicle routing problem," *Computers & Industrial Engineering*, vol. 120, pp. 360-381, 2018.

[12] Y. Anoshkina and F. Meisel, "Technician teaming and routing with service-, cost- and fairness-objectives," *Computers & Industrial Engineering*, vol. 135, pp. 868-880, 2019.

[13] A. Goel and F. Meisel, "Workforce routing and scheduling for electricity network maintenance with downtime minimization," *European Journal of Operational Research*, vol. 231, no. 1, pp. 210-228, 2013.

[14] S. Çakırgil, E. Yücel and G. Kuyzu, "An integrated solution approach for multi-objective, multi-skill workforce scheduling and routing problems," *Computers & Operations Research*, vol. 118, pp. 1-17, 2020.

[15] H. Allaham and D. Dalalah, "MILP of multitask scheduling of geographically distributed maintenance tasks," *International Journal of Industrial Engineering Computations*, vol. 13, no. 1, p. 119–134, 2022.

[16] M. Rastegar, H. Karimi and H. Vahdani, "Technicians scheduling and routing problem for elevators preventive maintenance," *Expert Systems with Applications*, vol. 235, pp. 1-20, 2024.

[17] M. Alzeraif, A. Cheaitou and A. Bou Nassif, "Predicting Maintenance Labor Productivity in Electricity Industry using Machine Learning: A Case Study and Evaluation," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 14, pp. 528-534, 2023.

[18] M. Fathian, M. Saei-Shahi and A. Makui, "A New Optimization Model for Reliable Team Formation Problem Considering Experts' Collaboration Network," *IEEE Transactions on Engineering Management*, vol. 64, no. 4, pp. 586-593, 2017.

[19] Google, "Google Maps," 2023. [Online]. Available: https://www.google.com/maps/@24.901993,52.5992552,2672618m/data=!3m1!1e3!5m1!1e1?entry=ttuG. [Accessed 10 September 2023].

Development of Crack Detection and Crack Length Calculation Method using Image Processing

Jewon Oh[1], Yutaka Matsumoto[2], Kohei Arai[3]
Dept. of Architecture-Faculty of Eng., Sojo University, Kumamoto City, Japan[1, 3]
Dept. of Architecture and Building Services Eng., Kurume Institute Technology, Kurume City, Japan[2]
Information Science Dept., Saga University, Saga City, Japan[3]

***Abstract*—To evaluate the integrity of a building, many experts and engineers have evaluated the damage classification of a building based on superficial visual information through field surveys. On-site surveys are hazardous and require several years of experience and expertise. In this study, a system for detecting the presence or absence of cracks and calculating their lengths was developed using image processing technology. The accuracy of the system was examined using crack image data obtained from shear force experiments. For crack detection, a crack detection method was developed using canny edge, threshold, and HSV color detection. The detection of the presence of cracks was proposed to be coupled with image segmentation to improve detection accuracy. A method for calculating the crack length using image processing was also developed. In this study, we proposed a method to calculate cracks as straight lengths, and obtained results with 98.1% accuracy. However, for curved cracks, it was necessary to rotate or segment the image.**

Keywords—Image processing; crack detection; length calculation; color detection; canny; threshold; OpenCV

I. Introduction

Once a building is constructed, it will last several decades before the next renovation or reconstruction [1]. Therefore, the "lock-in effect" is very large, which makes it difficult to address. Additionally, natural disasters and deterioration have reduced the functionality of many buildings [2]. To evaluate the integrity of a building, many experts and engineers have evaluated the damage classification of a building based on superficial visual information through field surveys [3, 4]. However, assessing damage classification from the superficial visual information of reinforced concrete (RC) requires many years of experience and expertise. Other field investigations involving hazards are difficult to conduct [5, 6]. In Japan, the Ministry of Land, Infrastructure, and Transport created an "Inspection Support Technology Performance Catalog" to provide inspection support technologies [7, 8]. However, the support is inadequate in terms of ensuring hazards.

In recent years, image processing technology has been used to determine building damage, which eliminates the need to conduct dangerous on-site surveys. Image processing techniques are evolving daily [9,10,11], and in the field of architecture, many studies have been conducted on crack detection using image classification techniques, such as tunnel crack detection [12,13], bridge and road crack detection [14,15,16], and fatigue degradation detection of material deterioration [17,18]. Many of these studies used existing trained models to detect the presence or absence of cracks [19, 20]. Crack detection has been studied for several years, and various detection methods have been proposed. However, the various local environmental factors make it difficult to respond efficiently. In addition, there is insufficient specific information regarding the expertise of engineers and experts. Therefore, methods for detecting cracks and evaluating damage classification have not been sufficiently investigated.

In this study, an experimental study was conducted on the shear strength of a reinforced concrete column with a wing wall on one side. [21]. The crack image data obtained in the experimental process were subjected to Canny edge detection [22], threshold detection [23], and HSV color detection [24] to investigate crack detection and crack length calculation methods. The accuracy of the model was verified by comparing it with visual crack detection. Finally, we aim to develop a system that can predict the future direction in which cracks will propagate and reach a failure mechanism. In this paper, we developed a method for detecting the presence or absence of cracks and a system for evaluating the degree of damage using image processing technology based on image data. In particular, an efficient crack-detection method using image segmentation is considered.

II. Literature Review

Research on crack detection using image processing techniques can be divided into civil engineering and architecture [25]. Civil engineering is primarily concerned with crack detection in tunnels, bridges, and roads. On the other hand, in the architectural, research has been conducted on crack detection caused by earthquake damage and deterioration.

First, we discuss previous research in the architectural field. Lu et al. [5] developed a method for determining the presence or absence of damage by Canny edge detection using image data of building exteriors. Wang et al. [26], Fan et al. [27] improved an existing crack detection model and conducted a case study on a data set to improve the model's performance. The shape of the cracks was confirmed by counting the pick cells in the crack image. Zhang et al. [28] proposed a crack detection model by adding an ensemble algorithm to existing crack detection models. Cracks could be detected in any type of image. Yamaguchi et al. [29, 30] proposed a crack detection method for concrete surfaces using image processing techniques. A method for efficiently detecting cracked areas in images using mask processing methods was identified.

Next, we discuss previous studies in the field of civil engineering will be presented. Liu et al. [31] used existing trained convolutional neural networks (CNN) models to classify cracks in buildings and roads. Cracks were detected differently depending on the number of pixels in the input image. Li et al. [32] attempted to use FoSA's learning genetic algorithms to detect cracks in roads. A case study was conducted using five image data sets. It was found to be difficult to detect depending on the brightness of the images. Wang et al. [33] used threshold detection in image processing technology to detect the degree of road damage at different brightness levels in the image. Kulambayev et al. [34] developed a method to detect the degree of road damage in real time using a deep learning model. Maeda et al [35] and Abbas et al. [36] installed cameras in cars and used CNN models to detect road cracks and abnormal conditions. Yadav et al [37] used three crack detection algorithms to detect crack presence/absence classification and crack patterns. 2-3% improved detection performance compared to existing crack detection systems.

Crack detection research is active in both architectural and civil engineering. However, most previous studies have focused on model improvement with new images and crack detection using trained models. Therefore, pretreatment methods for crack detection and crack length calculations have not yet been adequately studied. In this study, we developed a preprocessing method for crack images using image processing technology and clarified the method for calculating the crack length.

III. Methodology

In this study, image data of the cracks were collected through shear strength experiments. Subsequently, a crack presence/absence detection method and crack damage evaluation was conducted (see Fig. 1). Three full-scale 1/3-scale specimens were fabricated for shear strength experiments. This experiment was conducted to investigate the ultimate shear strength and failure mechanism of the columns with single-sided walls.

Crack image data were recorded on video with a 4 K camera installed in front of the experimental subject, as shown in Fig. 1. The video recording was performed continuously from the beginning to the end of the experiment. The video recording was set to 30 FPS at 3840 × 2160 pixels high quality and saved in *.MP4 files. The same settings were used to capture the videos of the three experimental subjects. The captured video data were converted into image data for crack detection using the following procedure (see Fig. 2). In addition, it was difficult to obtain the expected results in this study because each region in the image had a different brightness depending on the shooting situation. Therefore, we considered a method in which the input image was divided into 7 × 5 images, image processing was performed in more detail, and the images were merged into the original image.

1) The cracked video was edited only for the necessary parts. 3 hours (size:30.2 GB) video edited to 27 minutes (size:1.96 GB).

2) The edited video was saved as a still image by using OpenCV.

3) The still image was selected and cut out only where cracks occurred.

4) The cropped images were divided into 7 × 5 images.

5) Segmented images were used to detect the presence of cracks using image-processing technology.

6) The processed images were then combined and returned to the original image.

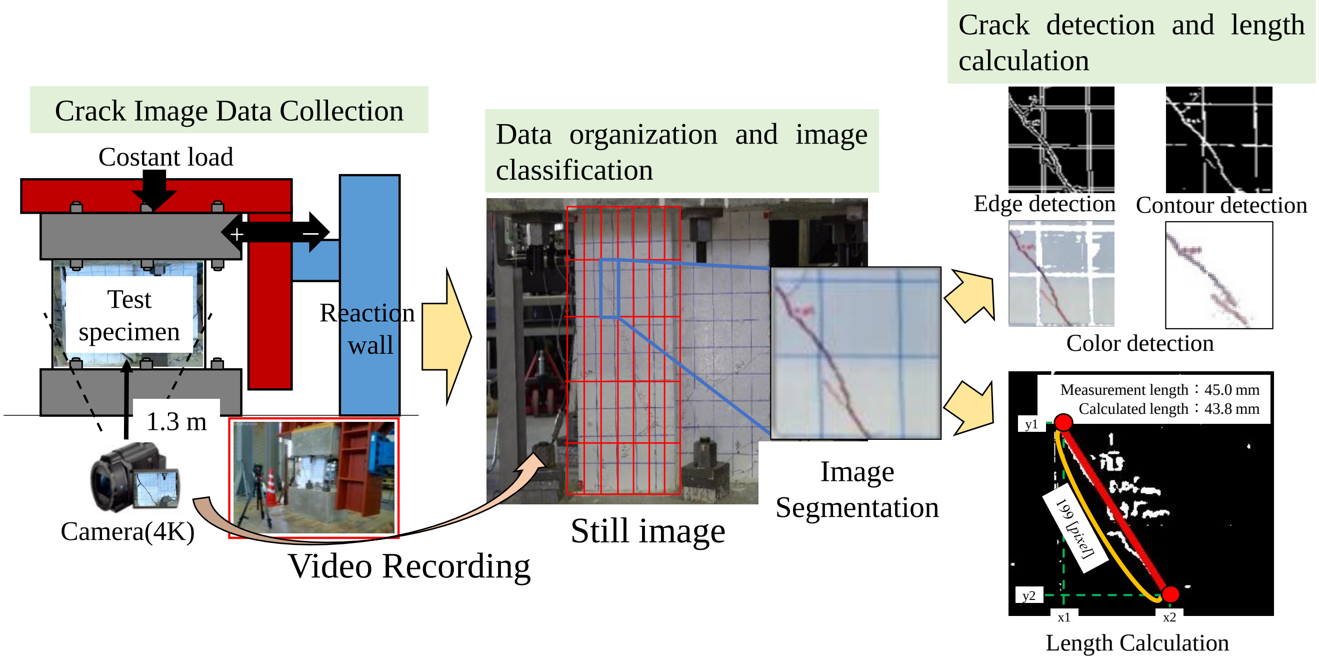


Fig. 1. An overview of this study.

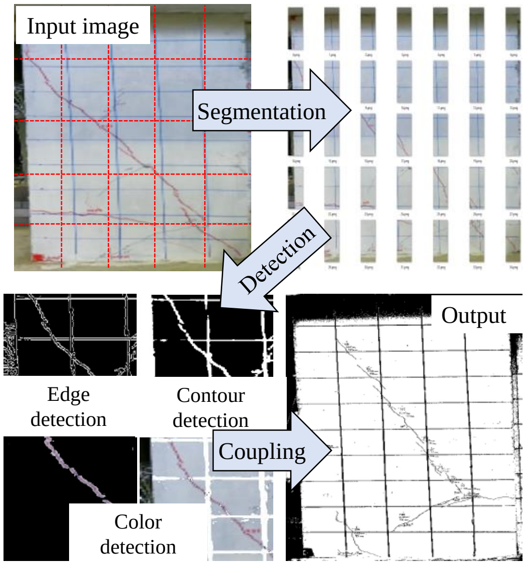


Fig. 2. Preprocessing for crack detection.

For crack detection, data organization and image processing were performed according to the above procedure.

IV. Experiment and Discussion

Crack detection was performed by organizing videos obtained from shear strength experiments and using image processing techniques to detect the presence of cracks. In this study, three image processing techniques were used to detect the presence of cracks. The three image processing techniques are Canny edge detection, Threshold detection, and HSV color detection. Each method for detecting the presence of cracks is described using the input images in Fig. 3. We also describe a method for calculating the length of cracks using image processing techniques.

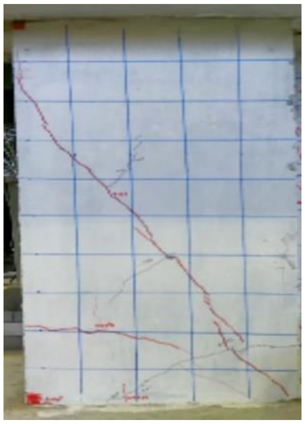
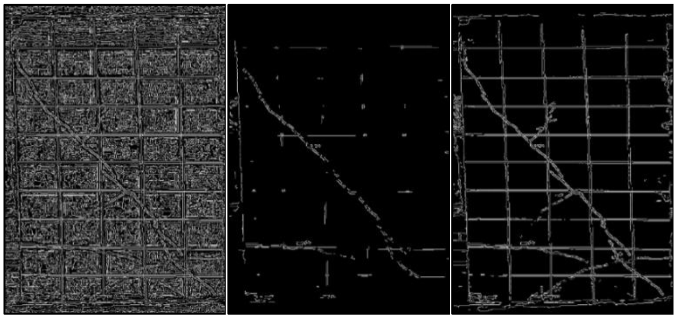


Fig. 3. Input image for crack detection.

A. Canny Edge Detection

Canny edge detection is a method for detecting vertical and horizontal edges that indicate points where the pixel values change abruptly in one direction [38]. The accuracy of edge detection depends on the value of the parameter. Therefore, it was necessary to adjust the values of the appropriate parameters. In this study, we employed OpenCV Track Bar to determine the appropriate parameter values. The Track Bar also allows for real-time fine-tuning of parameter values. Fig. 4 shows the results of crack presence detection by adjusting the parameter values. The smaller the value of the parameter, the more noise is generated and the more difficult it is to detect cracks (see Fig. 4(a)). However, the larger the value of the parameter, the less noise there is, but some cracks are removed along with the noise (see Fig. 4(b)). The results with appropriate parameter values indicated that the cracks were easily detected (see Fig. 4(c)).

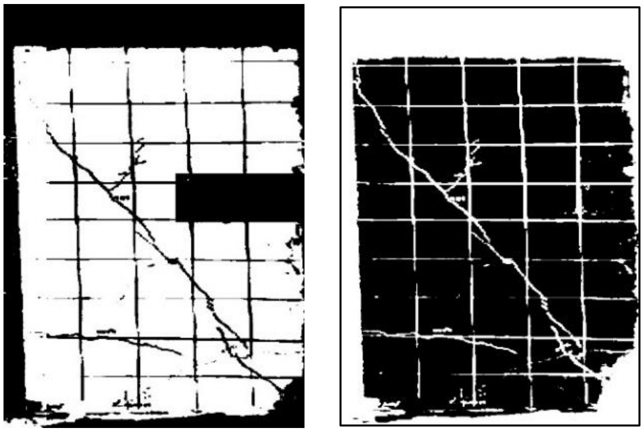


(a) Minimal parameters (b) Maximal parameters (c) Optimal parameters

Fig. 4. Edge detection using the Canny method

B. Threshold Detection

Threshold detection converts an image pixel value to white if it is greater than a parameter value and to black otherwise, and binarizes the image [39]. In this study, we used the THRESH_BINARY model, which adjusts parameter values directly, as in Canny edge detection, and the OTSU model, which adjusts them automatically [40, 41]. Both models were compared to examine their accuracy in detecting the presence of cracks. The THRESH_BINARY model uses a Track Bar to adjust the parameter values. Fig. 5 shows the results of the comparison of the models. Both models were able to detect cracks. However, the OTUS model automatically adjusts the values of the parameters; therefore, detection is not possible places. The OTUS model has a significant effect on the resolution of the input image.

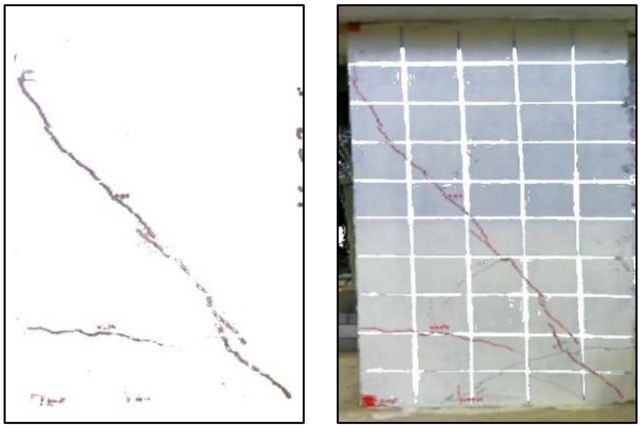


(a) OTUS (b) THRESH_BINARY

Fig. 5. Contour detection using the threshold method.

C. HSV Color Detection

In the shear force experiments conducted in this study, cracks were visually identified and marked during the experiment. Color markings were marked black for positive forces and red for negative forces. The authors considered a method to detect cracks by directly detecting color. In this study, we attempted to detect red and blue colors using the HSV color space model. Fig. 6 shows the results of the red and blue detection. In red detection, the method detects the areas marked in red and changes all other areas to white. In blue detection, the method detects blue and changes the color of the area to white. Both detection methods were able to detect cracks using color detection. However, some areas appear to have failed to be detected in certain regions. Color detection was found to be less accurate, depending on the resolution of the input image.



(a) Color detection (red) (b) Color detection (blue)

Fig. 6. Color detection results by HSV model.

D. Crack Length Detection

Crack lengths were calculated using OpenCV and Numpy. In this study, the crack lengths were calculated linearly. The crack length is determined by Eq. (1) to Eq. (4). Eq. (1) determines the number of pixels on the diagonal of an input image. In Eq. (2), the number of pixels is used to determine the resolution. Eq. (3) converts pixels to centimeters and calculates the length of one pixel. In Eq. (4), the length is calculated by counting the total number of pixels in the cracks to be calculated. The crack length was calculated by determining the values of the parameters based on the input image size. The value of parameter is the established reference point for the input image at the shooting distance (see Fig. 7).

$$dots \ = \sqrt{(w^2 \times l^2)} \tag{1}$$

$$dpi \ = dots \div inch \tag{2}$$

$$px \ = (25.4\, mm \div dpi) \times cv \tag{3}$$

$$L_t \ = px_t \times px \tag{4}$$

where $dots$ is the number of diagonal pixels of the input image [dots], w is the number of horizontal pixels of the input image [dots], l is the number of vertical pixels of the input image [dots], dpi is the resolution of the input image [dots/inch], $inch$ is the length of the diagonal of the input image [inch], px is the length per pixel of the input image [mm/pixel] (1 inch = 25.4 mm), cv is the correction factor for the input image size [mm], L_t is the total length between stickers [mm], and px_t is the total number of colored pixels [pixels].

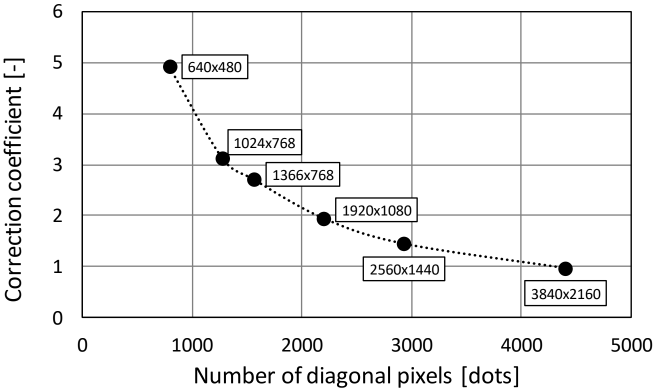


Fig. 7. Correction factor in image size.

In particular, Eq. (3) requires an accurate length per pixel. This length per pixel has a significant impact on the calculation accuracy. The accuracy of the length per pixel was determined using the blue border of the input image as the reference line. The blue line was 50 mm in length and width, which confirms the per-pixel validity. The following procedure was used to calculate the crack length (see Fig. 8).

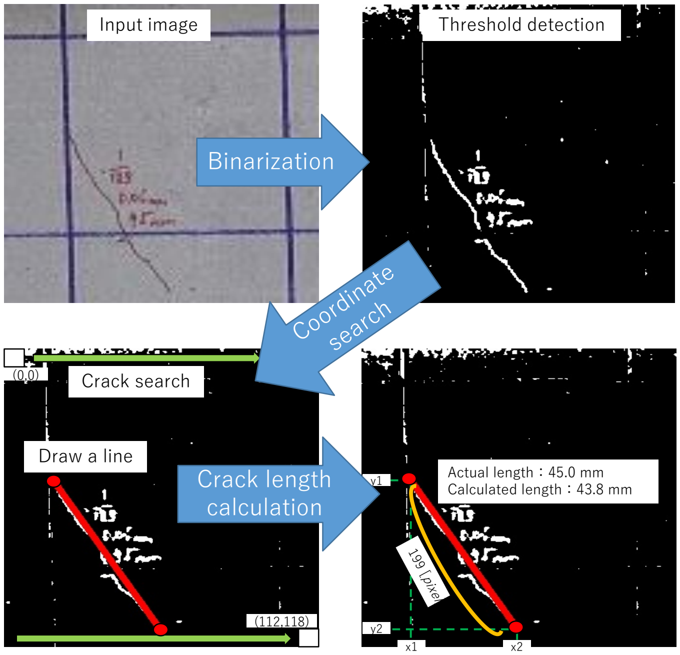


Fig. 8. Crack length calculation method.

1) The input image is converted into black and white binarizations.

2) Images converted to black and white were classified as the background and cracks, respectively.

3) The size of the input image was determined.

4) *Based* on its size, cracks are searched from the first to the last coordinates.

5) *Select* a location to calculate the length of the crack and draw a straight red line on the image.

6) *The* red pixels are counted linearly.

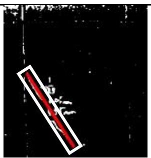
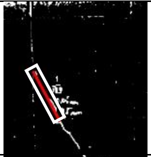
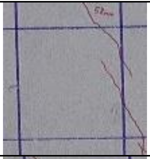
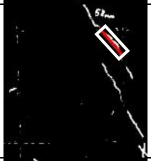
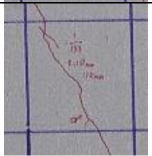
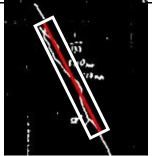
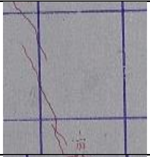
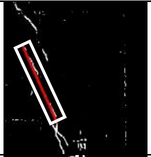
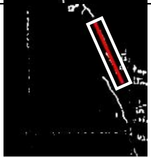
7) *The* crack length was calculated using Eq. (4).

E. *Crack Length Calculation*

Using the crack length calculation method described above, the lengths were calculated for the six cracks, as shown in Fig. 9. Table I lists the calculated lengths for each crack image. Images 1-6 were intended for cracks that were as continuous as possible. Cracks were examined through visual inspection during the shear force tests, and crack lengths were measured linearly using a tape measure. The measured length was compared to the calculated length to determine the accuracy of the calculation.

The results of the crack-length comparison yielded a 98.1% correct response rate. The closer the percentage of correct answers is to 100%, the more likely it is that the length is similar to the measured value. Crack lengths were calculated for crack images 3, 4, and 5, and were somewhat similar to the measured lengths for which crack lengths were compared. However, for images 1 and 2, the errors were 2.7% and 4.6%, respectively, which are much larger than the measurement results. It is thought that the point of the crack length measured at the time of measurement and the point of calculation were slightly different. It is also believed that the calculation error is affected by the accuracy of the binarization image processing. Other images of each crack had different image sizes depending on the area from which the crack was extracted. Therefore, the lengths corresponding to the pixels were different, which may have caused an error in the length calculation. In this study, the crack length was obtained linearly. If curved cracks are targeted, it is necessary to change the angle of the input image or to split the image.

TABLE I. COMPARISON OF ACTUAL CRACK LENGTHS AND CALCULATED RESULTS

No	Input image	Crack detection image	Image Size [pixel]	Measurements value [mm]	Calculated Value [mm]	Error rate [%]
1			112×118	45.0	43.8	2.7
2			112×118	28.0	26.7	4.6
3			107×107	15.0	15.1	0.7
4			107×108	40.0	40.2	0.5
5			134×111	56.0	56.6	1.1
6			113×105	30.0	29.5	1.7

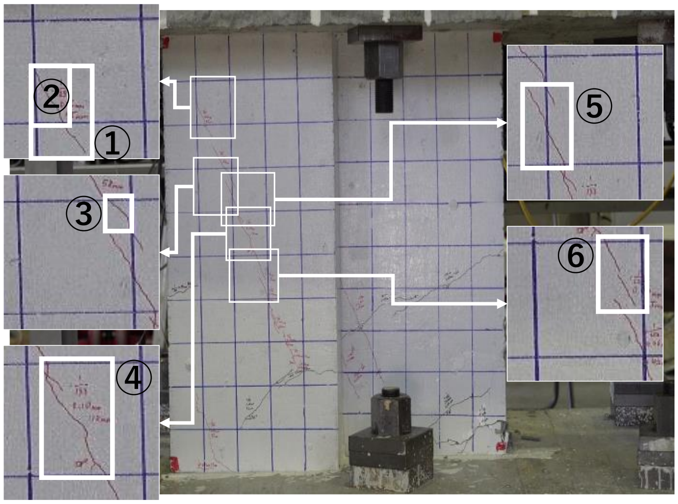


Fig. 9. Results for length of cracks.

V. Conclusion

In this study, crack image data were collected using a 4K camera during shear force experiments. Image data were used to develop a method for detecting the presence of cracks and calculating their lengths. Cracks were detected using edge, contour, and color detection. A method for calculating the crack length using image processing was also developed. The results obtained were as follows:

1) The visibility of the edge detection results differed depending on the adjustment of parameters.

2) The OpenCV Track Bar allows for easy parameter adjustment and proper crack detection.

3) The results obtained for contour detection varied significantly depending on the brightness of the image. As with edge detection, the accuracy of detection depends on the parameter values.

4) Color detection requires a different method for each color. In this paper, we propose a method that combines image segmentation.

5) The crack length calculation method is significantly affected by the length per pixel of the input image. Therefore, a reference line must be established and verified.

6) In this study, we proposed a method to calculate cracks as straight lengths, and obtained results with 98.1% accuracy. However, for curved cracks, it was necessary to rotate or segment the image.

7) For crack photography, adjusting the brightness using light according to the local environment is linked to detection accuracy.

8) A comprehensive study of crack detection methods tailored to these conditions is required to address a variety of environments.

Future Research Works

In the future, the machine learning of crack patterns will be used to build a model that can predict crack growth by classifying whether pattern deviations are within an acceptable range.

Acknowledgment

This research was supported by the President's Discretionary Fund of the Kurume Institute of Technology, an Obayashi Foundation Research Grant and JSPS KAKENHI Grant Number 22K14368. We express our gratitude to the reviewers.

References

[1] Alessio Mastrucci, Antonino Marvuglia, Ulrich Leopold, Enrico Benetto, Life Cycle Assessment of building stocks from urban to transnational scales:A review, Renewable and Sustainable Energy Reviews 74 (2017) 316–332, http://dx.doi.org/10.1016/ j.rser.2017.02.060.

[2] Janet L. Reyna, Mikhail V. Chester, The Growth of Urban Building Stock: Unintended Lock-in and Embedded Environmental Effects, Journal of Industrial Ecology, Volume 19, Issue 4 METHODS, TOOLS, AND SOFTWARE, pp.524-537, 2014, https://doi.org/10.1111/ jiec.12211.

[3] Ibrahim Baran Karasin, Comparative Analysis of the 2023 Pazarcık and Elbistan Earthquakes in Diyarbakır, Buildings 2023, 13, 2474. doi.org/10.3390/buildings13102474.

[4] Fatih Avcil, Investigation of Precast Reinforced Concrete Structures during the 6 February 2023 Türkiye Earthquakes, Sustainability 2023, 15, 14846. doi.org/ 10.3390/su152014846.

[5] Youcun Lu, Lin Duanmu, Zhiqiang (John) Zhai, Zongshan Wang, Application and improvement of Canny edge-detection algorithm for exterior wall hollowing detection using infrared thermal images, Energy & Buildings 274 (2022) 112421. doi.org/10.1016/j.enbuild.2022. 112421.

[6] Mohd Aliff, Nur Farah Hanisah, Muhammad Shafique Ashroff, Sallaudin Hassan, Development of Underwater Pipe Crack Detection System for Low-Cost Underwater Vehicle using Raspberry Pi and Canny Edge Detection Method, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, No. 11, 2022. DOI: 10.14569/IJACSA.2022.0131152.

[7] MLIT : A guideline for Function Continuity of Buildings that serve Disaster that Serve as Disaster Bases, 2019. (in Japanese). https://www.mlit.go.jp/ jutakukentiku/build/jutakukentiku_house_tk_ 000 088.html (accessed 2023.11).

[8] MLIT : Catalog of Inspection Support Technology Performance , 2022. (in Japanese). https://www.Mlit.go.jp/road/sisaku/inspection-support/ (accessed 2023.11).

[9] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Nature volume 521, pages436-444 (2015). doi:10.1038/nature14539.

[10] Hina Inam, Naeem Ul Islam, Muhammad Usman Akram, Fahim Ullah, Smart and Automated Infrastructure Management: A Deep Learning Approach for Crack Detection in Bridge Images, Sustainability 2023, 15, 1866. doi.org/10.3390/ su15031866.

[11] Niphat Claypo, Saichon Jaiyen, Anantaporn Hanskunatai, Inspection System for Glass Bottle Defect Classification based on Deep Neural Network, (IJACSA) International Journal of Advanced Computer Science and Applications Vol. 14, No. 7, 2023. DOI: 10.14569/IJACSA.2023.0140738.

[12] Yupeng Ren, Jisheng Huang, Zhiyou Hong, Wei Lu, Jun Yin, Lejun Zou, Xiaohua Shen, Image-based concrete crack detection in tunnels using deep fully convolutional networks, Construction and Building Materials 234 (2020) 117367. doi.org/10.1016/j.conbuildmat.2019. 117367.

[13] Tai-Tien Wang, Characterizing crack patterns on tunnel linings associated with shear deformation induced by instability of neighboring slopes, Engineering Geology 115 (2010) 80-95. doi:10.1016/j.enggeo. 2010.06.010.

[14] R.S. Adhikari, O. Moselhi, A. Bagchi, Image-based retrieval of concrete crack properties for bridge inspection, Automation in Construction 39 (2014) 180-194. doi.org/10.1016/j.autcon.2013.06.011.

[15] Md. Monirul Islam, Md. Belal Hossain, Md. Nasim Akhtar, Mohammad Ali Moni, Khondokar Fida Hasan, CNN Based on Transfer Learning

Models Using Data Augmentation and Transformation for Detection of Concrete Crack, Algorithms 2022, 15, 287. doi.org/10.3390/a15080287.

[16] Jinsong Zhu, Jinbo Song, An Intelligent Classification Model for Surface Defects on Cement Concrete Bridges, Appl. Sci. 2020, 10, 972. doi:10.3390/app10030972.

[17] Esraa Elhariri, Nashwa El-Bendary, Shereen A. Taie, Automated Pixel-Level Deep Crack Segmentation on Historical Surfaces Using U-Net Models, Algorithms 2022, 15, 281. doi.org/10.3390/a15080281.

[18] Dinar Mutiara Kusumo Nugraheni, Andi Kurniawan Nugroho, Diah Intan Kusumo Dewi, Beta Noranita, Deca Convolutional Layer Neural Network (DCL-NN) Method for Categorizing Concrete Cracks in Heritage Building, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 14, No. 1, 2023. DOI: 10.14569/IJACSA.2023.0140180.

[19] Baoju Liu, Tengyu Yang, Image analysis for detection of bugholes on concrete surface, Construction and Building Materials 137 (2017) 432-440. doi.org/10.1016/j.conbuildmat.2017.01.098.

[20] Luka Gradišar, Matevž Dolenc, Transfer and Unsupervised Learning: An Integrated Approach to Concrete Crack Image Analysis, Sustainability 2023, 15, 3653. doi.org/10.3390/su15043653.

[21] Yutaka Matsumoto, Shuichi Uehara, OH Jewon, Akihito Noguchi, Experimental Study on Shear Strength for Reinforced Concrete Column with Wing Wall on Either One Side, The 13th International Symposium on Architectural Interchanges in Asia 1233-1237, 2022.12.

[22] Zhao Xu, Xu Baojie, Wu Guoxin, Canny edge detection based on Open CV, 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), pp.53-56. DOI: 10.1109/ICEMI.2017.8265710.

[23] Fangyan Nie, Pingfeng Zhang, Jianqi Li, Dehong Ding, A novel generalized entropy and its application in image thresholding, Signal Processing 134 (2017) 23–34. http://dx.doi.org/10.1016/j.sigpro. 2016. 11.004.

[24] Manuel G. Forero, Julián Ávila-Navarro, Sergio Herrera-Rivera, New Method for Extreme Color Detection in Images, Springer Nature Switzerland AG 2020 K. M. Figueroa Mora et al. (Eds.): MCPR 2020, LNCS 12088, pp. 89-97, 2020. doi.org/10.1007/978-3-030-49076-8_9.

[25] Arun Mohan, Sumathi Poobal, Crack detection using image processing: A critical review and analysis, Alexandria Engineering Journal (2018) 57, 787-798. doi.org/10.1016/j.aej.2017.01.020.

[26] Wenjun Wang, Chao Su, Semi-supervised semantic segmentation network for surface crack detection, Automation in Construction 128 (2021) 103786. doi.org/10.1016/j.autcon.2021.103786.

[27] Zhun Fan, Chong Li, Ying Chen, Paola Di Mascio, Xiaopeng Chen, Guijie Zhu, Giuseppe Loprencipe, Ensemble of Deep Convolutional Neural Networks for Automatic Pavement Crack Detection and Measurement, Coatings 2020, 10, 152. doi:10.3390/coatings10020152.

[28] Xinxiang Zhang, Dinesh Rajan, Brett Story, Concrete crack detection using context-aware deep semantic segmentation network, Comput Aided Civ Inf. 2019;34:951-971. DOI: 10.1111/mice.12477.

[29] Tomoyuki Yamaguchi, Shingo Nakamura, Ryo Saegusa, Shuji Hashimoto, Image-Based Crack Detection for Real Concrete Surfaces, IEEJ Trans 2008; 3: 128-135. DOI:10.1002/tee.20244.

[30] Tomoyuki Yamaguchi, Shuji Hashimoto, Fast crack detection method for large-size concrete surface images using percolation-based image processing, Machine Vision and Applications (2010) 21:797-809. DOI 10.1007/s00138-009-0189-8.

[31] Yahui Liu, Jian Yao, Xiaohu Lu, Renping Xie, Li Li, DeepCrack: A deep hierarchical feature learning architecture for crack segmentation, Neurocomputing 338 (2019) 139-153. doi.org/10.1016/j.neucom.2019. 01.036.

[32] Qingquan Li, Qin Zou, Daqiang Zhang, Qingzhou Mao, FoSA: F* Seed-growing Approach for crack-line detection from pavement images☆, Image and Vision Computing 29 (2011) 861-872. doi:10.1016/j.imavis.2011.10.003.

[33] Penghui Wang, Yongbiao Hu, Yong Dai, and Mingrui Tian, Asphalt Pavement Pothole Detection and Segmentation Based on Wavelet Energy Field, Hindawi Mathematical Problems in Engineering Volume 2017, Article ID 1604130, 13 pages. doi.org/10.1155/2017/1604130.

[34] Bakhytzhan Kulambayev, Magzat Nurlybek, Gulnar Astaubayeva, Gulnara Tleuberdiyeva, Serik Zholdasbayev, Abdimukhan Tolep, Real-Time Road Surface Damage Detection Framework based on Mask R-CNN Model, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 14, No. 9, 2023. DOI: 10.14569/IJACSA.2023.0140979.

[35] Hiroya Maeda, Yoshihide Sekimoto, Toshikazu Seto, Takehiro Kashiyama, Hiroshi Omata, Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images, Computer-Aided Civil and Infrastructure Engineering 33 (2018) 1127-1141. DOI: 10.1111/mice.12387.

[36] Iman Hashim Abbas, Mohammed Qadir Ismael, Automated Pavement Distress Detection Using Image Processing Techniques, Engineering, Technology & Applied Science Research Vol. 11, No. 5, 2021, 7702-7708. doi.org/10.48084/etasr.4450.

[37] Dhirendra Prasad Yadav, Kamal Kishore, Ashish Gaur, Ankit Kumar, Kamred Udham Singh, Teekam Singh, Chetan Swarup, A Novel Multi-Scale Feature Fusion-Based 3SCNet for Building Crack Detection, Sustainability 2022, 14, 16179. doi.org/10.3390/su142316179.

[38] Huili Zhao, Guofeng Qin, Xingjian Wang, Improvement of Canny Algorithm Based on Pavement Edge Detection, 2010 3rd International Congress on Image and Signal Processing. DOI: 10.1109/CISP.2010.5646923.

[39] Shiping Zhu, Xi Xia, Qingrong Zhang, Kamel Belloulata, An Image Segmentation Algorithm in Image Processing Based on Threshold Segmentation, 2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System. DOI 10.1109/SITIS.2007.116.

[40] Yu-Kun Lai and Paul L. Rosin, Efficient Circular Thresholding, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 3, MARCH 2014. DOI: 10.1109/TIP.2013.2297014.

[41] Xiangyang Xu, Shengzhou Xu, Lianghai Jin, Enmin Song, Characteristic analysis of Otsu threshold and its applications, Pattern Recognition Letters 32 (2011), pp.956–961, doi:10.1016/j.patrec. 2011.01.021.

AUTHORS' PROFILE

Jewon Oh, He received BE, ME and PhD degrees in 2012, 2015 and 2021, respectively. He was with the appointed assistant professor at AI Application Laboratory, Kurume Institute of Technology in 2021. Lecturer at AI Application Laboratory, Kurume Institute of Technology in 2022. He is now an appointed assistant professor at Department of Architecture Faculty of Sojo University in 2023. His research is focused on developing energy-saving technologies in the building using AI and image processing.

Yutaka Matsumoto, He received Dr. degrees in 2018 respectively. From April 2001 to December 2005, he worked at Takada Corporation, and from January 2006 to March 2018 at S.A.I. Structural Design Co.,Ltd. From April 2018, he has been a professor at Kurume Institute of Technology, Department of Architecture and Building Services Engineering.

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January 1979 to March 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science in April 1990. He is now an Emeritus Professor of Saga University since 2014. He was a council member for the Aeronautics and Space related to the

Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of University of Arizona, USA since 1998 and is an Adjunct Professor of Nishi-Kyushu University as well as Kurume Institute of Technology/AI Application Laboratory since 2021. He also is Vice Chairman of the Science Commission “A” of ICSU/COSPAR since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 60 books and published 640 journal papers as well as 460 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html

Urban Image Segmentation in Media Integration Era Based on Improved Sparse Matrix Generation of Digital Image Processing

Dan Zheng*, Yan Xie
School of Humanities and Communication, University of Sanya,
Sanya, Hainan 572000, China

***Abstract*—Media integration integrates the resources of various media platforms, including audience, technical and human resources. In the era of media integration, media in various channels, at different levels and in different fields provide various choices for image communication and brand building of cities. The technology of image processing by computer has gradually affected all aspects of people's life and work, bringing more and more convenience to people. In this paper, the application of digital image processing technology in city image communication in the age of media integration is studied. A new sparse matrix creation method is proposed, and the created sparse matrix is used as the similarity matrix to segment the spectral clustering image, so that the edge contour weakened in gradient calculation can be corrected and strengthened again. The research shows that the improved algorithm is superior to the traditional algorithm, and compared with the fuzzy entropy algorithm based on exhaustive search, the gray contrast between regions and Bezdek partition coefficient are improved by 9.301% and 4.127%. In terms of speed, the algorithm in this paper has absolute advantages, so our research is also affirmed, which fully shows that it should have high application value.**

Keywords—Media integration; digital image processing; city image; image segmentation; improved sparse matrix generation

I. Introduction

Media combines the advantages of traditional media and new media, emphasizes the comprehensive application of text, pictures and image information, and increasingly highlights the value of images. By using computer technology, some textual descriptions are made into vivid images. Under the environment of media integration, the image of a city represents the characteristics of a city, and it is also the key factor that distinguishes it from other cities in the information explosion era. A good city image can not only show the achievements of urban construction, but also accelerate the process of urban development [1]. The technology of image processing by computer has gradually affected all aspects of people's life and work, bringing more and more convenience to people. In modern society, people's knowledge of unfamiliar cities is mainly obtained from the media, and the farther away the cognitive subject is from the city, the more people rely on the image spread by the media to get their knowledge, so the spread of city image cannot be separated from the media. On the platform of media, the new and old media penetrate, collide, advance and merge with each other, which constitute a prominent landscape of today's media ecology. Therefore, in the era of media integration, it is necessary to strengthen the extensive application of digital image technology, improve the production quality of publicity content, promote the steady improvement of information dissemination efficiency, and drive readers to form new information acquisition and reading habits.

For the analysis and research of city image, Mbaye and others think that the image and style of a city are the same. What the city style shows are a city's elegant demeanor, consciousness and landscape. It shows the cultural connotation and details of the city, the spiritual outlook of its citizens, and the development of its culture, economy, politics, science and technology [2]. M Ceren mainly divides the image of a city from two dimensions: subject and object. It is the main body of the city itself, that is, the overall style and features of the city, as well as its internal cultural connotation and quality, and the object is an abstract overall or intuitive evaluation of the city [3]. Kourtit et al. put forward the theory of urban image, first defined the concept of urban image, put forward five elements of urban image formation, and established a unique empirical research method of urban image [4]. Huijsmans et al., based on the theory of city image design and city management, discussed the potential and feasibility of city image visual management from the perspective of spiritual representation and management visualization [5]. Digital image processing technology mainly includes image processing, digital image coding, digital image inpainting, image segmentation and so on, which is widely used in different industries. In the field of media, digital image processing technology is mainly used in network transmission, advertising, film special effects production, post-processing and so on. According to recent published articles at home and abroad, the present situation and progress of digital image compression coding are carefully combed. Wang et al. summarized the basic theories and methods of digital image compression, and certified and studied several classical compression coding methods [6]. Da-Hai et al. adopted DCT (Discrete Cosine Transform), and its high bit rate (bit rate > 0.25 bit/pixel) can get a good compression ratio for continuous tone still gray or color images [7]. Ahmed et al. have made in-depth research on image segmentation methods based on graph theory, comprehensively utilizing various feature information, adopting various segmentation ideas and integrating various segmentation technologies, in order to reduce the computational complexity and improve the segmentation quality [8]. Wang et al. segmented the image by using the similarity of the internal feature information of the image area. However, the disadvantage of this method is that it is easy to generate false segmentation areas when segmenting complex and changeable natural images, thus resulting in over-segmentation [9]. Huang et al. proposed an image adaptive threshold selection algorithm based on wavelet analysis, which enables the feature points of image histogram to be expressed from coarse to fine by the feature points of wavelet transform, and enables the threshold to be adaptively selected [10]. Yan et al. proposed that the density and maximum distance product method should be used in the selection mechanism of the initial clustering center. User-interactive color migration is also proposed. This migration method selects sample blocks according to personal experience and objective basis, and specifies the corresponding relationship among the sample blocks, so as to realize color migration between images [11]. Oho et al. proposed an inpainting algorithm based on the interpolation of image gray level and gradient direction, which can simultaneously inpaint the topological structure and texture information in the image [12].

In the field of urban image segmentation, many researchers have proposed various methods based on sparse representation. However, these methods still suffer from high computational complexity and insufficient generalization ability when processing large-scale urban images. In addition, these methods often overlook features such as texture, color, and shape in urban images, resulting in poor segmentation performance. Therefore, it is necessary to improve the existing sparse representation methods to enhance the accuracy and efficiency of urban image segmentation.

In order to address the limitations and challenges of existing methods, this paper proposes a city image segmentation method in the media fusion era based on improved sparse matrix generation digital image processing. The main motivation of this method is to improve the accuracy and efficiency of urban image segmentation, while reducing computational complexity and data volume.

II. Research Method

A. Media Communication of City Image under the Environment of Media Integration

The public’s cognition of the formation of a city is mainly carried out through traveling, listening to other people’s stories, media reports, etc., and the public can’t personally perceive all aspects of the city because of objective factors such as distance. Therefore, the public will choose the mass media to understand the image of a city, and after constant publicity, they will have an overall impression of a city. A good city media image is conducive to shaping the city image and attracting more audiences, while a good city media image is conducive to shaping the city image. A good city image can only be shaped by media reports and dissemination, so it is of great significance to study the city media image to recognize the city image.

City image communication power is the ability to realize the effective communication of city image, which directly determines the audience’s awareness of city image. The knowledge about the city’s citizens’ contact with the outside public mainly depends on the mass media, but for the outside people in the city, such as tourists and migrant workers, they will not only get the city information through the mass media, but also supplement their cognition of the city image through the behavior identity of the city image. Through the planning of the city image communication power, the city image communication power can be brought into full play, thus enhancing the public’s awareness of the city image. In the process of city image communication, the choice of city image positioning is to select the image information that can occupy a certain position in the public’s mind and refine it from the public’s preference. With the public as the center, the image of the city can be more easily accepted by the public, thus gaining the public’s trust and goodwill. It provides information, channels and strategies for city image communication. At the same time, it deals with the feedback from the audience in a timely manner, and revises the communication strategies so as to better build the communication power of city image and establish a good city image service.

In the age of media integration, the media of various channels, different levels and different fields provide various choices for the image communication and brand building of the city. This kind of mass media’s inherent focusing and spreading effect, as well as its functions of contacting the society, monitoring the environment, inheriting civilization, entertaining the public and mobilizing the society, make the mass media strategy enlarged into the exclusive strategy of city image communication in many researches and practices. However, it is undeniable that traditional media still plays an important role today. Traditional media still has its unique advantages in spreading city image, such as strong human and material resources, long-term accumulated rich experience, and lack of rigor, profundity and sense of authority of network media. Unlimited diversification of information channels is the biggest law of communication in the age of media integration. Traditional media can’t monopolize the release of information. Self-media, represented by Weibo, has more and more right to speak, and has become an unquestionable and important part of information release channels.

In the era of media integration, the power of individuals and non-governmental organizations composed of individuals in the main body of city image communication cannot be ignored. The development and maturity of new media technology provide audiences with more platforms to spread information and express their personal views. In this highly open discourse space, ordinary citizens’ awareness of spreading the city image is rapidly awakened, and they gradually change from passive to active in spreading the city image. They begin to use various media to express the city image in their hearts. Citizens in the city register their Weibo accounts and share their life in Weibo. Perhaps it’s a city landscape, a city food, a feeling in city life, or a city building, all of which spread the image of the city in tangible and intangible ways. Because most people in the circle of friends are acquaintances, people will be more willing to share information about urban life on WeChat. The bits and pieces of urban life that people share on WeChat can be quickly seen by people in the circle of friends, which will lead to secondary transmission. Let the receiver of information interpret the image of the city, such as the urban buildings and customs contained in the published information. This kind of unconscious communication can sometimes bring unexpected communication effects.

At the same time, we should not only make use of emerging media to pay close attention to positive publicity, but also deal with negative information. The integrated media communication of city image must follow the communication law of multi-media era, comprehensively utilize all kinds of media, meet the information needs of different types of audiences, and carry out three-dimensional communication. In the way, we should not only keep the traditional advantages, but also be good at using new media. In terms of timing, we should not only have ideas and means in normal situations, but also take measures to deal with thinking and image restoration in critical times. City image communication has a distinct stage, changing with the times and its own development; from the microscopic point of view, any specific communication work is a certain stage in the process of city image communication. Compared with the other two forms of city image communication, interpersonal communication, memory communication and experience communication, the participation of the media will systematically and continuously spread the city image through news reports, which will have an impact on the audience's cognition and understanding of the city image in a long period.

B. *Analysis of the Application of Digital Image Processing Technology in City Image Communication*

With the advent of the age of media integration, images, as the visual basis for people to perceive the world, are an important means for people to obtain, express and transmit information. Computer technology is used to compress and encode images to complete digital image processing, and to transform low-quality images into high-quality images.

At present, the development of streaming media communication technology is relatively mature, and digital image processing technology is needed for image transmission. Digital image processing technology is widely used in streaming media, including high-speed color TV signal and high-compression image transmission. From the previous role of picture ornament and decoration to the application of digital image processing technology, more pictures with strong sense of scene, intuitive image and important subject matter are used, so that digital images can play a leading role in news reports.

1) Digital image compression technology: With the continuous development of multimedia technology and Internet technology, how to effectively organize, store, transmit and restore image data and explore more effective and higher compression ratio image coding technology has become one of the key tasks in information processing technology. the research and application of image compression coding is one of the most active fields in information technology at present. Image coding is to use different expression methods to reduce the amount of data needed to represent the image. the theoretical basis of compression is information theory. Without the rapid development of image compression technology, it will be difficult to realize the storage and transmission of large amount of information, and it will be difficult to play a role in multimedia and other application technologies. Therefore, data compression is very important for multimedia information such as images.

In order to express image data, some symbols need to be used, and image coding needs to use these symbols to express images according to certain rules [13-14]. Here, the symbol sequence assigned to each information or event is called the code word, and the number of symbols in each code word is called the length of the code word. Generally, the encoder includes three independent operations in sequence, while the corresponding decoder includes two independent operations in reverse order, as shown in Fig. 1:

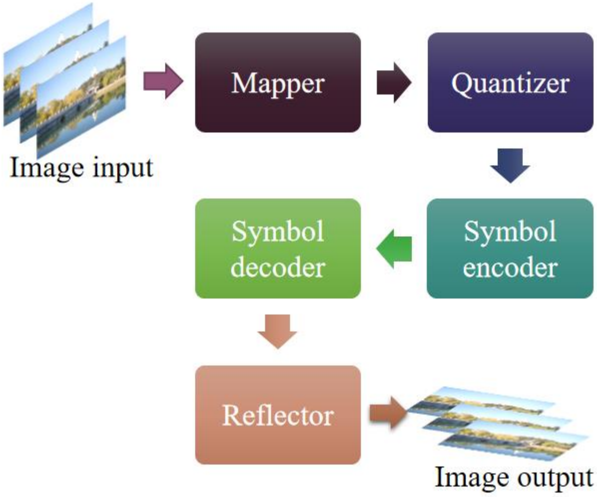


Fig. 1. Block diagram of image coding and decoding system.

In the encoder, the mapper reduces the pixel redundancy by transforming the input data; Quantizer reduces psychovisual redundancy by reducing the accuracy of mapper output; the encoder reduces coding redundancy by assigning the shortest code to the most frequent quantizer output value. Because the quantization operation is irreversible, there is no inverse operation module for quantization in the decoder [15].

For a random event E, if its occurrence probability is $P(E)$, then the information it contains is:

$$I(E)=\frac{1}{\log P(E)}=-\log P(E) \tag{1}$$

The source symbol set B is defined as the set (b_i) of all possible symbols, where each element b_i is called the source symbol, and the probability of the source generating the symbol (b_i) is $P(b_i)$.

For the effectiveness of image information transmission with large amount of data under the condition of high compression, the image to be compressed is preprocessed before it is compressed and coded to reduce the amount of information transmitted, and the decomposed wavelet coefficients can be appropriately changed, so that the compressed reconstructed image has better compression effect and subjective quality [16-17].

The “pure” anisotropic diffusion equation is adopted for the high frequency subband $HL_j, LH_j, HH_j \left(j = 1, \ldots, d \right)$. Carry out wavelet transform on the original image to obtain a wavelet image;

$$\left\{ u_{k,j} \middle| k = LL, LH, HL, HH;\ j = 1, \ldots, d \right\} \tag{2}$$

Initialization threshold T_0 (threshold to be used for the first scan):

$$T_0 = 2^{\left\lfloor \log_2 Max\left\{ \left| c_{i,j} \right| \right\} \right\rfloor} \tag{3}$$

Among them, $\left\{ c_{i,j} \right\}$ is the transform coefficient of L -level embedded zerotree wavelet transform, $\left| c_{i,j} \right|$ is the absolute value of $c_{i,j}$, and the relationship between the threshold of each scan and the threshold of its previous scan is: the current threshold is half of the previous threshold [18].

The sensitivity of human eyes to the change of gray scale is related to the background, and it changes with the change of average gray scale, that is, the resolution (vision) of human eyes to the details of the scene is related to the relative contrast C_r of the scene. The formula is as follows:

$$C_r = \left[\frac{B_1 - B}{B} \right] \tag{4}$$

where: B_1 is the brightness of the object; B is the background brightness; When C_r becomes small, vision decreases.

If a convolutional code with a code rate of R and a memory depth of M is used to send a message containing K information bits, the effective code rate is:

$$R_{eff} = \frac{K}{R^{-1}K + R^{-1}M} = \frac{RK}{K + M} = \frac{R}{\frac{M+1}{K}} \tag{5}$$

Therefore, convolutional codes are most effective at $K + M$, that is, the length of the information sequence to be sent is much longer than the storage length of the register.

2) Image segmentation technology: Image segmentation is an important image technology, which not only gets extensive attention and research, but also gets a lot of applications in practice. Image processing emphasizes the transformation between images to improve the visual effect of images. Image analysis is mainly to monitor and measure the objects of interest in the image, so as to obtain their objective information and establish a description of the image. They generally correspond to specific areas with unique properties in the image. In order to identify and analyze the targets, they need to be separated and extracted, and on this basis, it is possible to further utilize the targets. Image segmentation has been widely used in practice, such as industrial automation, online product inspection, production process control, document image processing, remote sensing and biomedical image analysis, security monitoring, military, sports, agricultural engineering and so on.

Segmentation is an important step in image analysis, image understanding and video coding, and it is also a basic technology in computer vision. This is because image segmentation, object separation, feature extraction and parameter measurement transform the original image into a more abstract and compact form, which makes it possible for higher-level analysis and understanding [19].

In this paper, the lost image information is analyzed, and a new sparse matrix creation method is proposed, and the created sparse matrix is used as similarity matrix for spectral clustering image segmentation. Gaussian function is used to calculate the similarity between pixels in an image.

$$w_{ij} = e^{\frac{-\left\| F(i) - F(j) \right\|_2^2}{\sigma_I}} * e^{\frac{-\left\| X(i) - X(j) \right\|_2^2}{\sigma_X}} \tag{6}$$

where: $F\left(\cdot\right)$ represents the gray value, and $X\left(\cdot\right)$ represents the coordinates. σ_X, σ_I represents the scale parameters of coordinate distance and gray difference, respectively.

At any point in a complex image, the human eye can only recognize dozens of gray levels, but it can recognize thousands of different colors. Therefore, in many cases, only using gray level information can’t extract a satisfactory target from the image. Color image segmentation can be regarded as the application of gray image segmentation technology in various color spaces.

Consider a two-dimensional gray (gradient) image I. The domain of I is $D_l \subset Z^2$, I takes discrete gray value $\left[0, N\right]$, which is regarded as the height of the corresponding pixel, and N is a positive integer:

$$I \begin{cases} D_l \subset Z^2 \to \left\{ 0, 1, \ldots, N \right\} \\ p \middle| \to I\left(p\right) \end{cases} \tag{7}$$

A path p of length l between points p, q in the image I is an $\left(l + 1\right)$ group composed of points $\left(p_0, \ldots, p_{l-1}, p_l\right)$, with $p_0 = p, p_l = q$ and $\forall_i \left[1, l\right], \left(p_{i-1}, p_i\right) \in G$.

In order to make the edge detected by edge detection operator affect the result of region segmentation, these edges should be mapped to the position of “dam” in gradient image.

Based on this, this paper puts forward a gradient "peak enhancement" method, which can correct and strengthen the weakened edge contour in gradient calculation. Based on the above analysis, the basic flow of this segmentation scheme is shown in Fig. 2.

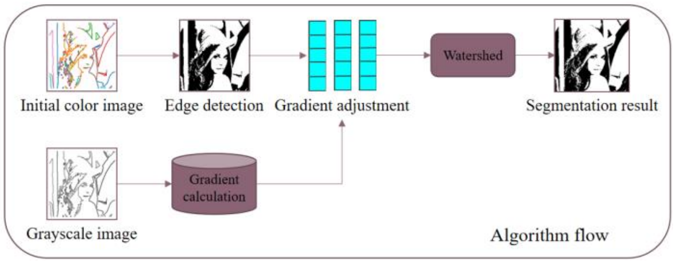


Fig. 2. Algorithm flow chart.

T is a threshold, and its value is related to the degree of noise pollution. It limits the range that the depth of point $\left(x,y\right)$ is considered to be very close to the depth of its neighboring points. Only when the depth value of point $\left(x,y\right)$ and its neighboring points are less than the threshold value, it is considered that point $\left(x,y\right)$ is very close to the depth value of its neighboring points.

$$T^2=\frac{1}{8}\sum_{i=0}^{7}\left(\delta_i\left(x,y\right)-f\left(x,y\right)\right)^2 \tag{8}$$

In order to make the color change independent of the change of light intensity in color image segmentation, an effective method is to uniformly obtain the change of intensity in spectral distribution. So a standardized color space is obtained, and its form is as follows:

$$r=\frac{R}{R+G+B} \tag{9}$$

$$g=\frac{G}{R+G+B} \tag{10}$$

$$b=\frac{B}{R+G+B} \tag{11}$$

Because of $r+g+b=1$, when two components are given, the third component is also determined. Usually, we only use two of the three components.

3) Image restoration technology: Image inpainting technology is a morbid problem. It is only a reasonable assumption based on mathematical principles, and it is analyzed from the perspective of computer vision and information theory. By comparing the rationality of various assumptions, the problem of image inpainting is finally solved. That is, the damaged area is artificially determined according to the color information and structural features in the image, and different colors are calibrated to distinguish the known area from the area to be repaired; Compared with the traditional manual method, today's image inpainting technology has developed by leaps and bounds in both efficiency and natural integrity. However, there are still a lot of problems in the field of image inpainting that need to be improved and solved urgently. As an important subject in image engineering, the field of image inpainting is attracting more and more researchers to join in.

Digital image inpainting technology is based on the problem that local information is missing in the process of digital image processing and compression in the past, and the object to be repaired is specially repaired. In particular, the restoration of cultural relics mainly involves manual restoration of cultural relics, which requires employees to have extremely high professional skills, patience and care, resulting in the inability of mass production of cultural relics restoration. However, the field of cultural relics restoration is very wide, and it is far from meeting the actual needs of cultural relics restoration only by manual restoration. For the possible negative effects of cultural relics restoration, digital image restoration technology can also provide certain anti-infection ability, avoid the serious impact of cultural relics restoration and improve the comprehensive utilization value of cultural relics [20]. Using the technical principles of statistics and other disciplines, the prediction model of cultural relics is constructed, and the existing image data is used to estimate the incomplete image area, so as to achieve the virtual restoration of cultural relics, which can shorten the original time-consuming work cycle of cultural relics restoration, and avoid the mistakes of manual restoration, resulting in more serious secondary damage to cultural relics. Reinforcement learning can be used for image segmentation tasks, learning how to classify images at the pixel level through interaction with the environment. The intelligent agent selects actions based on current observations and adjusts strategies based on feedback from the environment to optimize segmentation results. Reinforcement learning can be applied to target tracking tasks, where agents learn how to accurately track the position of target objects in consecutive frames. By defining appropriate reward functions, agents can learn to effectively track targets in complex scenarios. The combination of deep learning and object detection technology can achieve end-to-end object detection with higher accuracy and robustness. It trains neural networks to simultaneously predict the position and category of targets. For example, super-resolution reconstruction, style transfer, and image restoration. GAN consists of a generator and a discriminator, which generate new images similar to real images through adversarial training [21-23].

When studying image inpainting, the degradation model of the image can be expressed by the following formula:

$$u^0\big|_{\Omega\backslash D}=\left[k*u+n\right]\big|_{\Omega\backslash D} \tag{12}$$

Ω represents the target image; D represents the region to be repaired of the image; $\Omega\backslash D$ means the existing information in the image; u^0 is the available image part on $\Omega\backslash D$; u the target image to be restored; k represents its degradation function, n is the noise term, and "$*$" here represents convolution.

Generally, the energy form E for establishing its data model is defined by the minimum mean square error, and the formula is as follows:

$$E\left[u^0|u\right]=\frac{\lambda}{2}\int_{\Omega\setminus D}\left(k*u-u^0\right)^2dx \tag{13}$$

When partial differential equation theory is used for image inpainting, the D of the defect is unknown, so compared with other image processing projects, the prior model of the image is very important for the inpainting project.

Knowing the length, width and height of the cuboid can realize the complete reconstruction of the cuboid. The method proposed in this paper is to manually obtain the minimum number of image points, obtain their corresponding three-dimensional coordinates through camera calibration, and then calculate the length, width and height of the cuboid, thus realizing the complete reconstruction of the cuboid. The perspective projection of cuboid is shown in Fig. 3.

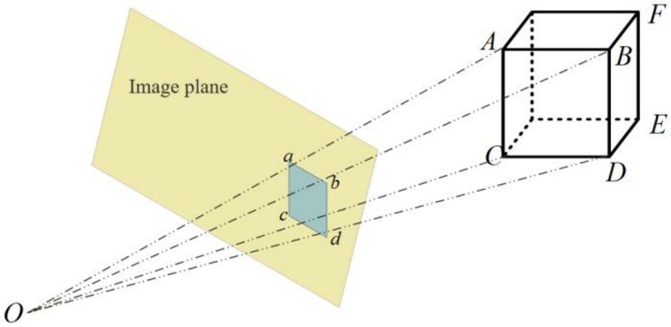


Fig. 3. Perspective projection of cuboid.

In the figure, O is the optical center of the camera, $ABCD$ is one surface of a cuboid, and its corresponding projection surface on the imaged image is $abcd$. Determine the depth value of one of the points in the camera coordinate system, calculate the three-dimensional coordinates of the manually obtained points according to the internal and external parameters calibrated by the camera, and then calculate the length, width and height of the cuboid according to its geometric properties.

The projection of the reverse $\vec{d}$ of any point x in the image is a ray determined by that point and the camera center. In the camera coordinate system, the ray direction $\vec{d}$ is:

$$\vec{d}=K^{-1}x \tag{14}$$

Therefore, A,B,C,D is on ray $\overrightarrow{Oa},\overrightarrow{Ob},\overrightarrow{Oc},\overrightarrow{Od}$.

In an image, the damaged area (the area to be repaired) is usually irregular, so it is difficult to determine the geometric center of the area. In this paper, the concept of gravity center in physics is introduced by using the characteristic that the position of gravity center is only related to the image shape.

Assuming that Ω represents the damaged area, the coordinates of its center of gravity O can be calculated by the following formula:

$$\bar{x}=\frac{\sum\sum_{[x,y]\subset\Omega}\|x\|}{S} \tag{15}$$

$$\bar{y}=\frac{\sum\sum_{[x,y]\subset\Omega}\|y\|}{S} \tag{16}$$

where, S represents the area of the region Ω to be repaired (the sum of all pixels to be repaired).

III. Analysis and Discussion of Results

In the basic fractal algorithm, the quality of reconstructed image, coding time and compression ratio depend on many factors, such as the size of R block, D block, the step length of generating D block pool, whether to adopt basic transformation, brightness adjustment factor and the quantization bit number of brightness offset, etc. the program is written in C++, the coordinates of D block are stored in 16bit, and the serial number of basic transformation is stored in 3 bits. The R block size is 4×4 and the D block size is 8×8(B=4), and eight basic transformations are adopted.

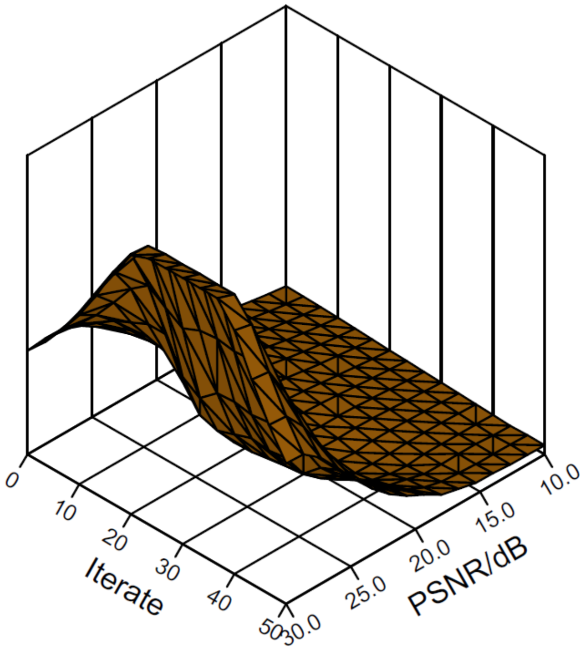


Fig. 4. The decoding convergence of lena image.

Analyze the decoding convergence of Lena image when the initial image is all black (see Fig. 4). It can be seen that the decoded image sequence of 10 times overlapping has basically converged. The difference between the decoded picture of the 10th overlapping and the PSNR (Peak Signal to Noise Ratio) after decoding is less than 0.051dB. Within the error range of 0.051dB, the decoded image of the 10th overlapping reaches the convergent image of the decoded image sequence.

Lena standard grayscale test chart with 516× 516, 8-bit quantization is used (see Table I and Fig. 5).

TABLE.I FRACTAL CODING

T	PSNR (dB)	Compression ratio	Coding time (s)
2	28.2349	5.8949	3.3796
6	26.1433	6.6526	2.8584
14	24.5524	12.3785	2.839

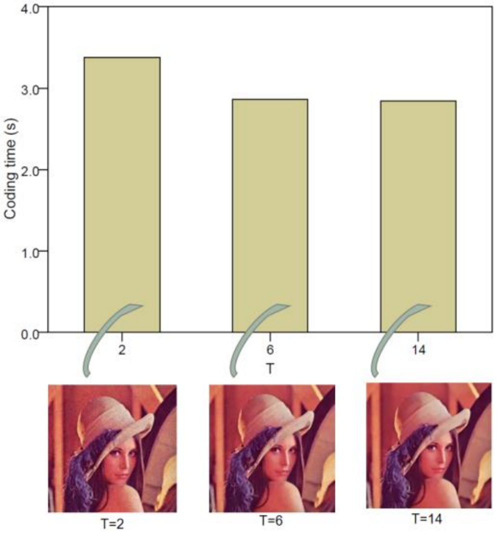


Fig. 5. Fractal time histogram.

The image reconstruction of neighborhood search fractal block coding is basically the same as that of global search fractal block coding, except that neighborhood search fractal block coding only iterates repeatedly in the neighborhood when reconstructing the image. This method limits the search process of the matching block to the four neighborhoods of the range block, which ensures the SNR and greatly reduces the coding time. However, with the increase of threshold, there is block effect in the decoded image, which is a shortcoming that this algorithm needs to overcome.

Influence of backtracking length on performance of convolutional codes. The length of backtracking not only determines the accuracy of Viterbi decoding, but also affects the decoding delay, and the error performance will also change accordingly. The feedback depth is 10, 15, 26 and 31 respectively. The simulation results are shown in Fig. 6.

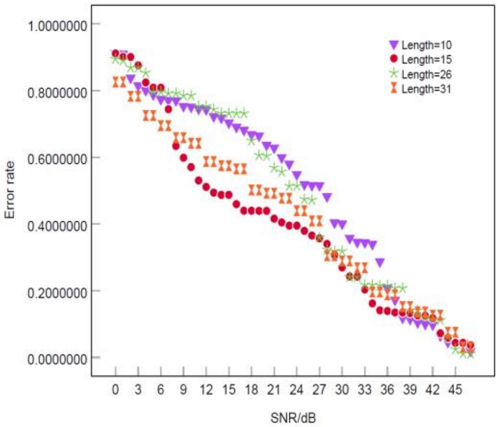


Fig. 6. Influence of different backtracking lengths on the performance of convolutional codes.

When the SNR (Signal to Noise Ratio) gradually increases, the bit error rate of the system gradually decreases with the increase of the backtracking length, but when the backtracking length increases to 5N, the bit error rate does not change much, and the (2, 1, 6) convolutional code here basically tends to be stable when the backtracking length reaches about 31. Therefore, when selecting the backtracking length, it is usually 5N.

Table II and Fig. 7 show the time taken by the fuzzy entropy algorithm based on exhaustive search and the algorithm in this paper to binarize sample images, where the unit of time is seconds.

The experimental results obtained by using the algorithm in this paper are basically the same as those obtained by fuzzy theory processing developed in the field of segmentation in recent years. However, the chart shows that the algorithm in this paper has absolute advantages in speed, so our research is also affirmed, which fully shows that it should have high application value.

To verify the effectiveness of the proposed algorithm, two quality indexes, i. e. gray contrast between regions and Bezdek partition coefficient, are selected to quantitatively analyze and compare the three algorithms involved. In general, the greater the gray contrast between regions, the better the image segmentation quality.

TABLE.II COMPARISON OF TIME

Image sequence number	Fuzzy entropy	Our
1	176.4963	0.6217
2	76.4064	1.1349
3	75.695	0.1376
4	185.0856	0.3185
5	183.5306	0.9046
6	229.1597	0.9303
7	91.1172	0.9817
8	193.1467	0.3634
9	202.0152	0.4158
10	172.9026	0.5851
11	205.0024	1.0143
12	55.7282	1.0713
13	159.9875	0.4138
14	174.3161	1.0269
15	64.1679	0.971
16	105.5269	0.2388
17	85.0808	1.0401
18	70.3764	0.0164
19	143.932	0.6305
20	196.2099	0.5157

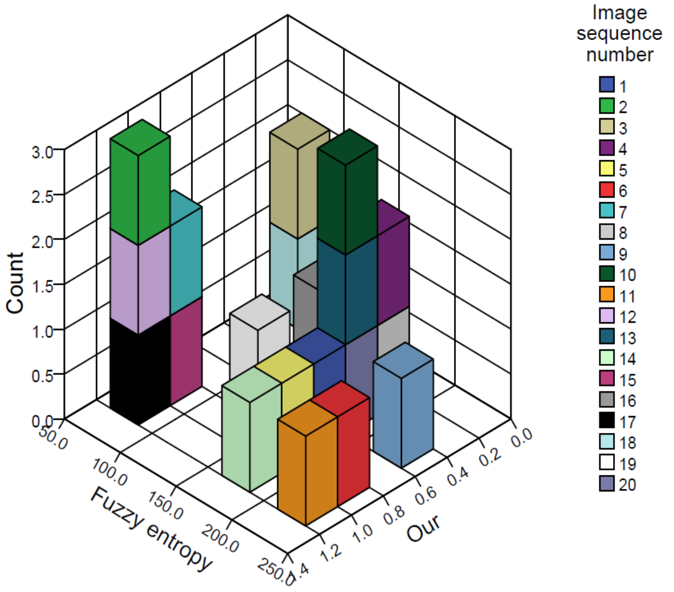


Fig. 7. Algorithm time comparison chart.

If there are multiple areas in a picture, calculate the contrast between two adjacent areas and then sum and compare. The larger the Bezdek partition coefficient is, the greater the membership of pixels within a class is, and the smaller the membership of pixels between classes is, and the better the segmentation effect is. The experimental results are shown in Table III and Table IV, Fig. 8 and Fig. 9.

TABLE.III COMPARISON OF GRAY CONTRAST BETWEEN REGIONS

Image sequence number	Fuzzy entropy	Peak detection	Our
1	0.1725	0.1974	0.4955
2	0.1973	0.2026	0.4995
3	0.2063	0.2082	0.5029
4	0.2117	0.2395	0.53
5	0.2362	0.2577	0.5546
6	0.2507	0.2675	0.5554
7	0.2527	0.2821	0.5599
8	0.2672	0.2875	0.6057
9	0.3004	0.3153	0.6216
10	0.303	0.3321	0.6252
11	0.3313	0.3391	0.6429
12	0.3357	0.3479	0.6494
13	0.3358	0.3483	0.6544
14	0.3595	0.3577	0.6575
15	0.3666	0.3666	0.7828
16	0.3989	0.375	0.7864
17	0.4099	0.4208	0.7965
18	0.4297	0.4475	0.8678
19	0.466	0.4687	0.8799
20	0.4743	0.4968	0.8824

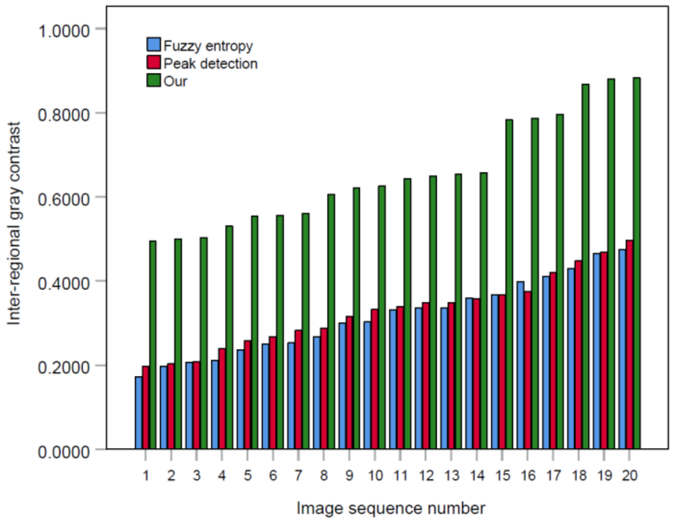


Fig. 8. Comparison chart of gray contrast between regions.

TABLE.IV COMPARISON OF BEZDEK PARTITION COEFFICIENTS

Image sequence number	Fuzzy entropy	Peak detection	Our
1	0.8723	0.8344	0.9953
2	0.8575	0.8332	0.9794
3	0.852	0.827	0.974
4	0.85	0.823	0.9732
5	0.8438	0.8185	0.9599
6	0.8417	0.8123	0.9586
7	0.8416	0.8048	0.9557
8	0.8376	0.7924	0.952
9	0.8282	0.789	0.9394
10	0.8158	0.7869	0.9348
11	0.8157	0.7853	0.9345
12	0.8143	0.7734	0.9306
13	0.8134	0.7616	0.9238
14	0.8089	0.7548	0.9176
15	0.8033	0.7499	0.9122
16	0.8021	0.7259	0.9059
17	0.7911	0.7256	0.8953
18	0.7907	0.689	0.8851
19	0.7768	0.6832	0.88
20	0.7606	0.6703	0.862

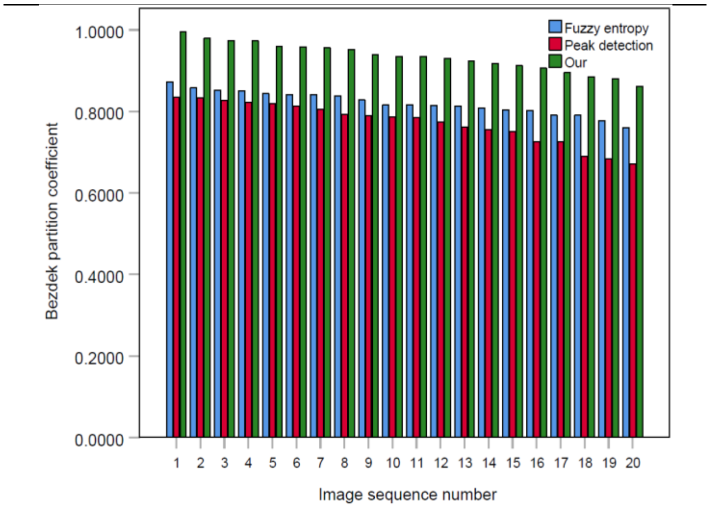


Fig. 9. Comparison chart of Bezdek partition coefficient.

In this paper, the original image is pre-segmented by using the established graph model. The initial clustering center of the algorithm is obtained by hierarchical clustering of the obtained similarity matrix. It can be seen from the two quality indexes of inter-regional gray contrast and Bezdek partition coefficient involved in the above chart that the improved algorithm in this paper is superior to the traditional algorithm. Compared with the fuzzy entropy algorithm based on exhaustive search, the inter-regional gray contrast and Bezdek partition coefficient are improved by 9.301% and 4.127%.

The urban image segmentation method based on improved sparse matrix generation for digital image processing in the era of media fusion has broad application prospects in the real world. For example, in urban planning and construction, this method can help urban managers better understand and analyze urban spatial layout and landscape characteristics, providing more accurate data support for urban planning decisions. At the same time, this method can also be applied to urban image dissemination strategies, by using precise image segmentation techniques to highlight the characteristics and highlights of the city, and improve its visibility and reputation.

In addition, the urban image segmentation method in the media fusion era based on improved sparse matrix generation digital image processing can also be applied in fields such as tourism, transportation, and environmental monitoring. For example, in the field of tourism, this method can help tourists better understand the characteristics and landscape layout of tourist attractions, and improve the quality of tourism experience; In the field of transportation, this method can help traffic management departments better understand road conditions and traffic flow, providing more accurate data support for traffic planning and governance; In the field of environmental monitoring, this method can help environmental protection departments better understand the situation of environmental pollution and ecological change trends, and provide more accurate data support for environmental protection decision-making.

IV. Conclusion

City image communication power is the ability to realize the effective communication of city image, which directly determines the audience’s awareness of city image. On the platform of media, the new and old media penetrate, collide, advance and merge with each other, which constitute a prominent landscape of today’s media ecology. Through the planning of the city image communication power, the city image communication power can be brought into full play, thus enhancing the public’s awareness of the city image. In this paper, the application of digital image processing technology in city image communication in the age of media integration is studied. A new sparse matrix creation method is proposed, and the created sparse matrix is used as similarity matrix to segment spectral clustering images. Gaussian function is used to calculate the similarity between pixels in an image. the research shows that the improved algorithm is superior to the traditional algorithm, and compared with the fuzzy entropy algorithm based on exhaustive search, the gray contrast between regions and Bezdek partition coefficient are improved by 9.301% and 4.127%.

When implementing image segmentation methods based on improved sparse matrix generation for digital image processing, there may be the following potential limitations and challenges: sparse representation and feature extraction processes involve a large number of matrix operations and optimization problems, with high computational complexity. Urban images usually contain a large amount of pixel information, and the processing process requires a large amount of data, which requires high computational resources and storage capacity. For urban images under different scenes and lighting conditions, this method may not achieve ideal segmentation results. In order to overcome the potential limitations and challenges mentioned above, future research can focus on more efficient sparse representation and feature extraction algorithms, reducing computational complexity and time costs using deep learning techniques to learn and extract features from urban images, improving segmentation accuracy and generalization ability.

Acknowledgments

This work is supported by Hainan Provincial Philosophy and Social Sciences Planning Project: “Research on the governance and guidance of public opinion on the prevention and control of major epidemics in Hainan” (NO. HNSK(YB)20-50); the program of Teaching Reform of the Education Department of Hainan Province “Research on the reform of Media Literacy education for Adolescents in Hainan” (NO. Hnjg2021-97).

References

[1] Zhou, X., Zhang, X., Dai, Z., Hermaputi, R. L., & Li, Y. (2021). Spatial layout and coupling of urban cultural relics: analyzing historical sites and commercial facilities in district iii of shaoxing. Sustainability, 13(12), 6877.

[2] Mbaye, J., & Dinardi, C. (2019). Ins and outs of the cultural polis: informality, culture and governance in the global south. Urban Studies, 56(3), 578-593.

[3] M Ceren. (2018). Book review: urban music and entrepreneurship: beats, rhymes and young people’s enterprisewhitejoyurban music and entrepreneurship: beats, rhymes and young people’s enterpriseroutledge, london, 2017, ￡110.00 hbk (isbn: 9781138195462), 160 pp. Cultural Sociology, 12(1), 116-118.

[4] Kourtit, Nijkamp, & Romo. (2019). Cultural heritage appraisal by visitors to global cities: the use of social media and urban analytics in urban buzz research. Sustainability, 11(12), 3470.

[5] Huijsmans, T., Harteveld, E., Brug, W., & Lancee, B. (2021). Are cities ever more cosmopolitan? studying trends in urban-rural divergence of cultural attitudes. Political Geography, 86(2), 102353.

[6] Wang, Q., Mao, X., Jiang, X., Pei, D., & Shao, X. (2021). Digital image processing technology under backpropagation neural network and k-means clustering algorithm on nitrogen utilization rate of chinese cabbages. PLoS ONE, 16(3), 23.

[7] Da-Hai, Xia, S., Song, L., Tao, Z., Qin, Z., & Wu, Z. (2020). Review-material degradation assessed by digital image processing: fundamentals, progresses, and challenges. Journal of Materials Science & Technology, 53(18), 148-164.

[8] Ahmed, A. S., & Omar, A. (2018). Detecting incipient radial deformations of power transformer windings using polar plot and digital image processing. IET Science Measurement Technology, 12(4), 492-499.

[9] Wang, D., Zhao, Y., Rong, L., Wan, M., Shi, X., & Wang, Y. (2019). Expanding the field-of-view and profile measurement of covered objects in continuous-wave terahertz reflective digital holography. Optical Engineering, 58(2), 2-7.

[10] Huang, Y. F., & Wu, H. Y. (2020). Image retrieval based on asift features in a hadoop clustered system. IET Image Processing, 14(1), 138-146.

[11] Yan, Z., Zhang, H., Wang, X., M Gaňová, Lednick, T., & Zhu, H. (2022). An image-to-answer algorithm for fully automated digital pcr image processing. Lab on a Chip, 22(7), 1333-1343.

[12] Oho, E., Suzuki, K., & Yamazaki, S. (2020). Applying fast scanning method coupled with digital image processing technology as standard acquisition mode for scanning electron microscopy. Scanning, 2020(6), 1-9.

[13] Khatri, A. (2018). Assessment of road surface condition using digital image processing. Journal of Engineering Technology, 6(2), 612-625.

[14] Du, Z., Yuan, J., Xiao, F., & Hettiarachchi, C. (2021). Application of image technology on pavement distress detection: a review. Measurement, 184(3), 109900.

[15] Chouhan, S. S., Koul, A., & Singh, U. P. (2018). Image segmentation using computational intelligence techniques: review. Archives of Computational Methods in Engineering, (2), 1-64.

[16] Gibril, M., Idrees, M. O., Yao, K., & Shafri, H. (2018). Integrative image segmentation optimization and machine learning approach for high quality land-use and land-cover mapping using multisource remote sensing data. Journal of Applied Remote Sensing, 12(1), 1.

[17] Tinkler-Davies, B., & Shah, D. U. (2021). Digital image correlation analysis of laminated bamboo under transverse compression. Materials Letters, 283(7), 128883.

[18] D’Haen, J., May, M., Knoll, O., Kerscher, S., & Hiermaier, S. (2021). Strain acquisition framework and novel bending evaluation formulation for compression-loaded composites using digital image correlation. Materials, 14(20), 5931-.

[19] Karthikeyan, N., Saravanakumar, N. M., & Sivakumar, M. (2021). Fast and efficient lossless encoder in image compression with low computation and low memory. IET Image Processing, 15(11), 2494-2507.

[20] Belda, R., R Megías, Feito, N., A Vercher-Martínez, & Giner, E. (2020). Some practical considerations for compression failure characterization of open-cell polyurethane foams using digital image correlation. Sensors, 20(15), 4141.

[21] Valente, J., António, J., Mora, C., & Jardim, S. (2023). Developments in Image Processing Using Deep Learning and Reinforcement Learning. Journal of Imaging, 9(10), 207.

[22] Wang, X., Sun, L., Chehri, A., & Song, Y. (2023). A Review of GAN-Based Super-Resolution Reconstruction for Optical Remote Sensing Images. Remote Sensing, 15(20), 5062.

[23] Jiang, X., Wu, Z., Han, S., Yan, H., Zhou, B., & Li, J. (2023). A multi-scale approach to detecting standing dead trees in UAV RGB images based on improved faster R-CNN. Plos one, 18(2), e0281084.

Towards a Stacking Ensemble Model for Predicting Diabetes Mellitus using Combination of Machine Learning Techniques

Abdulaziz A Alzubaidi, Sami M Halawani, Mutasem Jarrah
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract—**Diabetes Mellitus (DM) is a chronic disease affecting the world's population, it causes long-term issues such as kidney failure, blindness, and heart disease, hurting one's quality of life. Diagnosing diabetes mellitus in an early stage is a challenge and a decisive decision for medical experts, as delay in diagnosis leads to complications in controlling the progression of the disease. Therefore, this research aims to develop a novel stacking ensemble model to predict diabetes mellitus a combination of machine learning models, where an ensemble of Prediction classifiers was used, such as Random Forest (RF), Logistic Regression (LR), as base learners' models, and the Extreme gradient Boosting model (XGBoost) as a Meta-Learner model. The results indicated that our proposed stacking model can predict diabetes mellitus with 83% accuracy on Pima dataset and 97% with DPD dataset. In conclusion, our proposed model can be used to build a diagnostic application for diabetes mellitus, as recommend testing our model on a huge and diverse dataset to obtain more accurate results.**

Keywords—DM; Diabetes Mellitus; Stacking; Ensemble learning; Machine Learning; Random Forest (RF); Logistic Regression (LR); Extreme Gradient Boosting model (XGBoost)

I. Introduction

Diabetes Mellitus mainly leads to chronic hyperglycemia considering low insulin quantities in the bloodstream [1]. Insulin plays an important role in glucose level lowering in the blood, carbohydrate anabolism, physical growth, cell reproduction, and protein and fat anabolic statute [2]. Therefore, severe difficulties are associated explicitly with Diabetes Mellitus concerning people's quality of life. The chronic diseases caused by Diabetes Mellitus have heart failure, kidney failure, blindness, and cardiovascular illnesses [3]. These conditions induce a high pile in mortality rates and pressures on personal life [4]. According to estimates, there was 463 million people worldwide with diabetes in 2019 [5]. Moreover, by 2030, that number is expected to rise to 578 million and then 700 million (2045).

Urban areas (10.8% percent) have a higher prevalence than rural areas (7.2% percent), and high-income countries (10.4% percent) have a higher prevalence than low-income ones (4.0% percent) [5]. 50.1% percent of people with diabetes don't know they have the disease.

According to estimates, there is 7.5% percent (374 million) people worldwide who have impaired glucose tolerance, and that number is expected to rise to 8.0% percent (454 million) by 2030 and 8.6% percent (548 million) by 2045 [5]. There are two prevalent forms of DM: Type-1-Diabetes (T1DM), an autoimmune syndrome that results in the death of beta cells in the pancreas that produce insulin, and Type-2-Diabetes (T2DM), which is a chronic condition that frequently results in abnormally high blood sugar levels (glucose) [6]. T1DM affects about 10% percent of patients under 30, whereas T2DM affects about 90% percent of diabetics over 30% percent [6]. Doctors use trial results from agreed-upon studies to distinguish between these types and then specify the best treatment options based on the form they have discovered. Medical experts occasionally disagree on the proper type of diagnosis, which makes treating the illness challenging [7].

Diabetes is becoming more prevalent worldwide, particularly in middle-income nations [8]. Therefore, we need to conduct this study to predict diabetes using machine learning methods to support doctors in providing the most suitable treatment strategy.

It was noted during the literature reviews that the emphasis is on ensemble machine learning techniques for predicting diabetes mellitus therapy and prevention are challenging due to suitable policies to provide environments that support healthy behaviors and a lack of quality health care in various settings.

The Sustainable Development Community seeks to eliminate premature mortality for various illnesses, including diabetes, by 2030 [8]. As a result, experts are continually researching multiple facets of DM where many machine learning techniques are used such as the RF model, which is a great choice in binary classification processes, for example in diabetes mellitus, the outcome is that the patient is diabetic or not diabetic, where the random forest depends on ensemble learning method (bagging) in making the final decision [33], [32]. The LR algorithm also plays an important role in the process of predicting diabetes, as it identifies the independent variables and classifies them on the x- and y-lines, and then measures the probabilities of an event's [34],[31]. One of the recently discovered options for machine learning is the XGBoost model which also counts on ensemble learning methodology; it can deal with unbalanced datasets classes by measuring the loss function and resolve the problem of overlearning using grading and voting [27], [28], [29].

A medical diagnosis of diabetes mellitus is one of the challenges in the medical field. Patients' information may include age, body mass, triceps skinfold thickness, serum insulin, plasma glucose concentration, diastolic blood pressure, and other factors. Based on these elements, the decision will made. The decision-making process is drawn out and takes weeks or months, making the doctor's job incredibly hard, but with the help of new technologies, it will be easier; consequently, machine learning techniques are a crucial solution [9].

Today, an extensive selection of medical datasets that are helpful for research in fields of medical science are easily accessible [10]. According to all this information and background about diabetes mellitus disease and the most prominent techniques used to predict it, we propose a novel stacking ensemble model for the prediction of DM utilizing a combination of machine learning models.

The Contributions of this paper are as follows:

- Developing a stacking ensemble model for predicting Diabetes Mellitus using a combination of machine learning models.
- Merging the RF and LR models as base learners and the XGB model as a meta-learner in building the proposed stacking model.

This paper is arranged as follows: Section II provides the Related work; Section III covers the material and methodologies of our study; Section IV illustrates the experimental setup for our proposed model; and Section V the performance measures. Our results and discussion are discussed in Section VI and VII respectively. Finally, the Conclusion has been addressed in Section VIII.

II. Related Work

Ensemble learning is a computational and statistical approach. Mimicking how people learn social skills by experimenting with different viewpoints before making the final decision. A Set of machine learning models combines choices and provides more robust and accurate predictions [11-12].

Gollapalli et al. [13] proposed a novel stacking ensemble model using machine learning to detect three-types of diabetes mellitus: T1DM, T2DM, and Pre-diabetes. Empirical results showed that the proposed model could predict with 94.48 percent accuracy, 94.48 percent recall, 94.70 percent precision, and 0.917 percent Cohen's kappa score. After observation, the most critical features of predicting T1DM, T2DM are: Sex; human gender A1c: measures the amount of sugar bonded to the hemoglobin protein in the blood; TG: The blood triglyceride level of the patient; LDL: Low-density lipoprotein, or LDL, is a measure of the quantity of harmful cholesterol; AntiDiab: A blood sugar-lowering oral medicine used to combat diabetes; Albumin: The amount of protein; Insulin, Injectable, Nutrition, Education. However, the study needed more ML classifiers and deep learning models to increase prediction accuracy.

Dutta et al. [14] emphasize that using an ML-based ensemble model in predicting DM is critical in ensuring more accurate predictions. Also, exploring deep learning techniques and applying them with an ensemble learning approach is recommended. Stacking is an ensemble method that employs a meta-model in which a novel classifier integrates multiple weak learners to predict the target variable [13].

Ganie and Malik [15] discussed the various ensemble learning methods, such as the Bagging method, in predicting T2DM based on lifestyle indicators. The synthetic minority oversampling technique is used for dataset class balancing. Furthermore, the results are validated using the Cross-Validation technique. Researchers and practitioners use the cross-validation technique for the model-building process to remove biases.

Laila et al. [16] studied efficient ensemble algorithms for predicting diabetic risks in the early stages, using Seventeen features gathered from the UCI of various datasets. This research used predictive models like (AdaBoost, Bagging, and Random Forest) were utilized to evaluate accuracy, recollection, and F1-score. Overall, the RF ensemble methodology had the highest accuracy (97 %), whereas AdaBoost and Bagging had lesser accuracy.

Javale and Desai [17] concentrated on an ensemble technique for healthcare information analytics employing machine learning through unbalanced dataset approaches, synthetic minority over-sampling, plus adaptive synthetic over-sampling. Using other analysis techniques such as the train-test, the K-folds, and the repeat train-test. The average Stacking-C technique was used to execute an ensemble strategy on the diabetes dataset, which included K-Nearest Neighbors (KNN), Support vector machines (SVM), RF, Naïve Bayes (NB), and logistic regression classifiers. The Synthetic Minority Oversampling Technique (SMOTE) reduces False Negative counts with more precision. An ensemble method facilitates appropriate decision-making by providing a more profound knowledge of the implementation. Rather than just comparing the classifiers' outputs produced for various performance measures, choosing the optimal ensemble technique for the application is always preferable. The fundamental challenge in healthcare information analytics is unbalanced datasets, which might be a critical factor for an ensemble technique in healthcare data analytics.

Singh et al. [18] suggested an ensemble-based approach for diabetes prediction called eDiaPredictTo forecast diabetes status in patients, it employs ensemble modeling, which consists of an ensemble of multiple machine learning algorithms such as XGBoost, RF, SVM, NN, and DT. The minimalizing error value and lowest weighted coefficient of eDiaPredict have all been tested. The suggested approach's usefulness is shown using the PIMA Indian medical dataset, which has an accuracy of 95% percent. The stacking ensemble combines the predictions of many ML models to get the maximum accuracy achievable compared to the conventional models. It leverages a single model known as a meta-model to diagnose the optimum mix of expectations from the basic models. The stacking ensemble contains two stages, level 0 and level 1. The former employs heterogeneous ML models known as base learners. In contrast, the latter uses a single model known as a meta learner, whose purpose is to unify the predictions of the basis learners. To predict T2DM and alert patients in advance to decrease the risk factor and intensity associated with diabetic diseases.

Geetha and Prasad [19] suggested a hybrid ensemble model. For the decision tree, they employed ensemble approaches such as bagging with "random forest" and Adaboost and supervised classification algorithms like Naive Bayes. Merge different two or more models improves performance by increasing the accuracy and precision of predictions. Joshi et al. [31] focused on predicting Type 2 diabetes in Pima Indian women using a logistic regression (LR) model and a decision tree, and the accuracy of the proposed model was 78% percent.

Patil et al. [20] proposed a framework for T2DM prediction that uses a stacking-based ensemble with a "non-dominated sorting genetic algorithm" method. The main objective is to reduce the time elapsed between diabetes diagnosis and medical evaluation. The suggested NSGA-II stacking method was compared to Boosting, Bagging, RF, and Random Sub-space approaches. The stacking ensemble methodology has outperformed all other traditional ensemble approaches. Findings indicate that the NSGA-II stacking approach performs better over other conventional ensemble methods with an accuracy of 81 percent.

Syed and Khan [21] created an ML-Based System for Predicting the Risk of (T2DM), which is a web-based prediction model that uses Azure ML to estimate the risks of Type 2 diabetes. The results show that the suggested model can accurately predict the risks of Type 2 diabetes by 82 percent. The geographical range of this study was restricted since it primarily focused on the western portion of Saudi Arabia for the validation procedure. Table I explains the most important studies according to limitations and Advantages, Data Sources.

III. Material and Methodologies

This research proposes a stacking ensemble model to predict diabetes mellitus. The proposed model relies on two essential levels of construction. The first level is called (base learners). At this level, a combination of machine learning models is prepared, trained, and produced predictions that are entered as inputs to a new model/classifier that learns from these inputs to make the final prediction (the meta-learner), the second level. We have selected the logistic regression model, Random Forest, as base learners with their distinction ability in binary classification processes. We select The XGBoost model as a meta-learner, which contributes positively to dealing with imbalanced dataset classes by minimizing the loss function and increasing the weight of the classified incorrect classes. The optimization GridSearchCV technology is applied to get the best possible results by the base learners and the meta learner; it uses a grid search of hyperparameters tuning for each model and extracts the best results. In our proposed model, we have included the cross-validation technique with default five-fold iterations to get optimal results. Also, we applied this technique on each of the base learners: Random Forest, Logistic regression, via the Optimizer GridSearchCV to get the best results by using the sci-kit-learn library, which provides a random split into training and test sets can easily calculate with the train_test_split assistant function. Each model starts with using K-1 of the folds. Fig. 1 describes the methodology for our proposed stacking model:

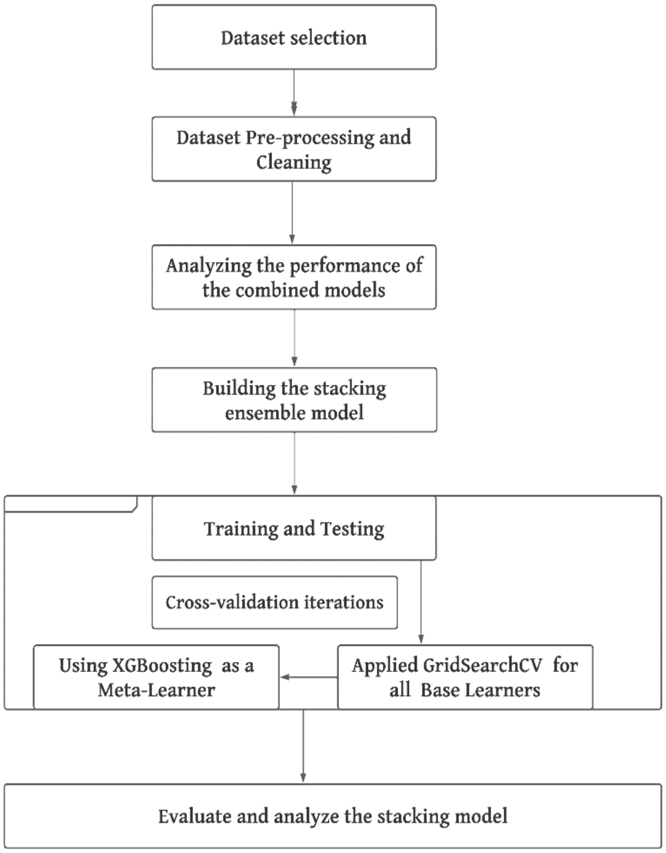

Dataset selection
Dataset Pre-processing and Cleaning
Analyzing the performance of the combined models
Building the stacking ensemble model
Training and Testing
Cross-validation iterations
Using XGBoosting as a Meta-Learner
Applied GridSearchCV for all Base Learners
Evaluate and analyze the stacking model

Fig. 1. Proposed stacking model.

A. Stacking

Stacking is an ensemble learning method that uses a meta-model where a new classifier integrates several individual-based learning predictions to get the best combined predictions. The stacking method has two levels in the building; level 0 (base-learners) combine heterogeneous models that are fitted and trained on a dataset, then the results will be fed as input to the meta-learner at the next level. The level 1 (meta learner) learns how to combine the predictions from the base models and provide robust and high-accuracy predictions [13]. We built our proposed model based on this stacking ensemble method with more of our contributions, like utilizing of cross-validation technique for all learners and leveraging the GridSearchCV hyperparameters tuning technique for the base learners. Fig. 2 illustrates the ensemble stacking methodology. Where m*n means that n of number k-folds Cross-validations of training dataset that will go cross all base learners' models, and m*M means that m of numbers of predictions coming from a number of M base learners will send to the next meta-learner model as inputs and then he learns from all of these predictions how to predict the final prediction.

TABLE I. IMPORTANT STUDIES IN PREDICTING DIABETES MELLITUS

Ref.	Algorithms	Data Sources	Advantages	Limitations
M. Gollapalli et al, 2022	Support Vector Machine, K-nearest Neighbor and Decision Tree.	Hospital (KFUH). Saudi Arabia	Use of Cross-validation technique in training the models, which leads to increased performance in prediction.	Need for using more and different machine learning models to improve results.
A. Dutta et al, 2022	Decision tree, Random Forest, Extreme Gradient Boosting, Light gradient boosting machine.	DDC dataset Bangladesh.	Use of Hyperparameter Optimization (Grid Search) for tuning the models.	Need for a large dataset to improve results.
A. Singh et al, 2021	Extreme Gradient. Boosting, Random. Forest, Support Vector Machine, Neural Network, and Decision tree.	PIMA Indian diabetes	Use of Recursive Feature Elimination (RFE) for feature space reduction in the dataset	Application of the proposed model in medical life tests.
A. H. Syed and T. Khan, 2020	Decision forest.	PIMA Indian diabetes.	Use of SMOTE technique for balancing dataset classes. which leads to avoiding overfitting.	Geographical scope of the study.
S. M. Ganie and M. B. Malik, 2022	Bagged Decision Trees, Random Forest, Extra Trees, AdaBoost, Stochastic Gradient Boosting, Voting.	Manually	Using the seaborn-Facet Grid method to visualize the dataset elements.	Develop an application for the proposed model to predict type 2 diabetes.
S. Härner and D. Ekman (2022)	Decision tree, Naïve Bayes.	PIMA Indian diabetes.	Comparing ensemble methods for predicting diabetes mellitus.	Need for using hyperparameters search optimizer to improve model results.

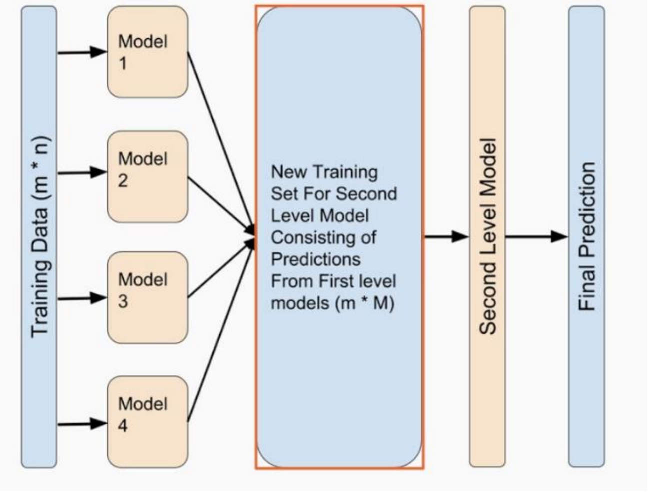

Training Data (m * n)
Model 1
Model 2
Model 3
Model 4
New Training Set For Second Level Model Consisting of Predictions From First level models (m * M)
Second Level Model
Final Prediction

Fig. 2. Stacking ensemble method.

B. GridSearchCV

GridSearchCV is a widely used technique in machine learning and deep learning model development. It helps select the best hyperparameter values for a specific model. Hyperparameters determine the model's behavior and configuration, such as the number of layers and batch size in neural networks or the depth of trees in decision trees.

GridSearchCV works by systematically testing different combinations of hyperparameters and evaluating their performance through cross-validation. This involves defining a set of possible values for each hyperparameter, training and testing the model with each combination, and ultimately selecting the combination that yields the best performance [14]. This process enhances the model's performance and prevents overfitting by evaluating it on separate training and testing data sets [14]. This approach significantly enhances model performance and mitigates overfitting by utilizing cross-validation, assessing the model on distinct training and testing datasets. [14].

C. Logistic Regression

Logistic regression is a machine learning algorithm used for binary classification. It is specifically designed to build models that can predict specific classifications. Despite its name, it is not used for predicting logistic events, but rather for classification based on the logistic (sigmoid) function [23].

Logistic regression solves the binary classification problem by classifying examples into one of two classes (e.g., class 0 or class 1) based on a set of independent variables (features). The algorithm uses the logistic function to model and determine the probability of an example belonging to the positive class (class 1) [24], [25]. In logistic regression, the results of the logistic function are transformed into probabilities using the sigmoid function. This conversion ensures that the probabilities range between 0 and 1. These probabilities represent the estimated classification probability, which is used to make the final classification decision [31]. Logistic regression is widely used across various fields, including data science, business analytics, medicine, and text classification. It is valued for its simplicity and its ability to handle large datasets efficiently [23].

D. Random Forest

Random Forest is a machine learning technique used for classification and prediction. It is a significant algorithm in optimization and diversification for classification and prediction models [32].

Random Forest is based on a collection of Decision Trees, a decision Tree divides data into categories by making sequential choices [32]. Each choice splits the data into subsets based on specific questions about the available variables. Random Forest creates a set of Decision Trees randomly by:

1) Randomly selecting samples with replacement from the original dataset (training data) for each decision tree.

2) Randomly selecting a subset of variables to build each tree.

Once the set of Decision Trees is constructed, Random Forest combines the individual tree predictions through voting or averaging to make a final decision for classification or prediction. Random Forest offers advantages such as model diversity and reducing overfitting, which occurs when the model becomes overly specialized to the training data. It also uses variable importance information to assess the impact of each variable on classification, providing valuable insights into the data [32]. Random Forest is widely used in various applications including image classification, word recognition, price prediction, and environmental analysis [32].

E. Extreme Gradient-Boosting

The gradient-boosting decision tree (GBDT) is the foundation of XGBoost, which was proposed by Tianqi Chen et al. [26]. A gradient-boosting algorithm built on a decision tree is called GBDT. Gradient boosting is an ensemble learning method that combines several weak classifiers into a stronger classifier during training. The objective of computing negative gradients is to enhance the following training cycle by minimizing the loss function and increasing the weight of the classified incorrectly classes. In contrast to GBDT, XGBoost incorporates a regularization technique to minimize model complexity, improve loss function smoothness, and prevent overfitting. To improve gradient boosting, locate the best-split solution, and promote scalability and efficiency, an approximation approach is also applied. XGBoost additionally enables parallel operations and an early stop to speed up the model operation. The model tree can stop to speed up training when the forecast result reaches the optimum. The model's classification accuracy can also be increased with XGBoost [27]. Zhao et al. [28] stated that XGBoost could effectively prevent the training model from over-fitting. Secondly, embedded parallel processing allows a faster learning speed.

Moreover, the XGBoost classifier can learn from imbalanced training data by setting class weight and taking ROC as evaluation criteria. XGBoost is one of the best classifiers for dealing with imbalanced datasets when the dataset classes with less variance [29]. Consequently, in this research, we chose it as the meta-learner of our proposed stacking model. The Extreme Gradient Boosting (XGBoost) is a machine learning algorithm used for classification and prediction tasks. It is an evolution of gradient boosting, combining multiple simple models into a strong and accurate model to enhance performance and accuracy.

XGBoost creates a sequence of weak models, like shallow decision trees, and boosts their performance. This boosting process focuses on the data predicted incorrectly by previous models, improving the overall performance of the model. Key features of XGBoost include:

1) Performance enhancement: XGBoost is known for achieving superior performance in various classification and prediction domains.

2) Multi-objective versatility: It can be used for both classification problems and numerical value prediction.

3) Overfitting prevention: Boosting parameters can be adjusted to limit overfitting and prevent excessive learning from training data.

4) Time and resource optimization: XGBoost strikes a balance between speed and accuracy, optimizing performance and resource utilization.

5) Handling missing data: XGBoost intelligently handles missing values without extensive preprocessing.

Overall, XGBoost is a powerful and popular tool in machine learning. It can effectively solve complex problems and improve predictive model performance.

F. Cross-Validation

Cross-validation is a popular data resampling method for evaluating a predictive model's generalization capacity and preventing overfitting by splitting the dataset into k-folds during training and testing iterations. The term "fold" here describes the quantity of generated subsets. The learning set's cases are randomly sampled for this division without being replaced. k-1 subsets comprise the training set and are used to train the model. The quality of this technology lies in storing unseen data at each new n-fold, which makes the prediction result more accurate. The model's performance is evaluated after being applied to the final subset, the "unseen dataset." This process is repeated until each of the k subsets has acted as a validation set [22]. Fig. 3 illustrates the cross-validation technique [30]. To achieve the best results, we used the cross-validation technique with the default four-fold iterations in our stacking model. We also used this technique on each of the base learners: RF, LR, and the Meta-Learner XGBoost via the Optimizer GridSearchCV to achieve the best results by using the sci-kit-learn library, which provides a random split into training and test sets that can be easily calculated with the train_test_split assistant function.

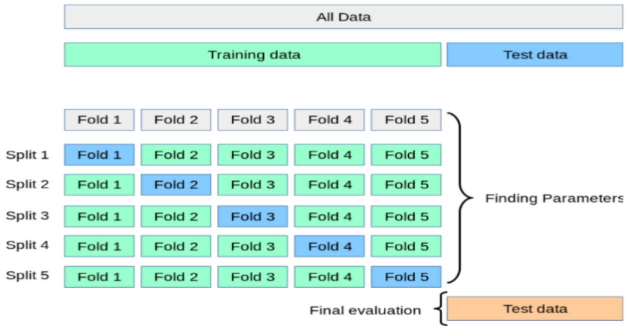


Fig. 3. Cross-Validation technique.

G. Study Dataset

The PIMA dataset, also known as the PIMA Indian Diabetes dataset, is a well-known dataset used in machine learning and data mining. It is named after the Pima Native American tribe in Arizona, USA. This dataset is commonly used for classifying the onset of diabetes in individuals by using medical diagnostic measurements [13]. It's available on Kaggle worldwide datasets repository, link: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.

This dataset consisted of 268 diabetic (positive = 1) and 500 non-diabetic (negative = 0) patients with eight Features presented below and in Table II. [13]:

1) Pregnancies: Number of pregnancies.

2) Glucose: Plasma glucose concentration measured two hours after an oral glucose tolerance test.

3) Blood Pressure: Diastolic blood pressure (mm Hg).

4) Skin Thickness: Triceps skinfold thickness (mm).

5) Insulin: Serum insulin level measured two hours after consumption (mu U/ml).

6) BMI: Body mass index (weight in kg / (height in meters) ^2).

7) Diabetes Pedigree Function: A function that estimates the likelihood of diabetes based on family history.

8) Age: Age in years.

TABLE II. PIMA DATASET INFORMATION

Data columns (total 9 columns):			
#	Column	Non-Null Count	Datatype
0	Pregnancies	768 non-null	int64
1	Glucose	768 non-null	int64
2	Blood Pressure	768 non-null	int64
3	Skin Thickness	768 non-null	int64
4	Insulin	768 non-null	int64
5	BMI	768 non-null	float64
6	Diabetes Pedigree Function	768 non-null	float64
7	Age	768 non-null	int64
8	Outcome	768 non-null	int64

Pima dataset is commonly used to demonstrate various machine learning techniques, such as logistic regression, decision trees, support vector machines, and neural networks, for predicting the likelihood of diabetes based on these medical measurements. It's important to note that while the PIMA dataset is valuable for educational purposes and experimenting with machine learning algorithms, it is relatively small and has limitations, such as missing values and potential biases. Therefore, caution should be exercised when drawing conclusions or developing predictive models solely based on this dataset in real-world applications.

Statistical analysis provides essential tools for visualizing and understanding the dataset pattern to improve the data pre-processing and modeling process. Fig. 5 presents the statistical information of the features and data types found in the PIMA dataset. Table III shows the distribution of the features based on count, mean, standard deviation, maximum value, minimum value, and the percentile/quartile of each feature. The correlation coefficient has been used to measure the feature relationships in Fig. 6, and finally, outcomes values are presented in Fig. 4.

IV. EXPERIMENTAL SETUP

In this research, we used Jupyter Notebook to build the stacking ensemble model, using Microsoft Intel(R) Core i5-1035G7 CPU 1.20GHz and 8 Giga RAM. The public Pima Dataset has been selected, pre-processed, and cleaned up from a few defects, such as zero values in features columns, using the arithmetic mean for each column. The base learners' models were initialized with a Random Forest model using the Grid-search Hyperparamets Tunner through the following Hyperparameters: 'bootstrap training,' 'max of samples training,' 'max_features,' 'min_samples_leaf,' 'min_samples_split,' 'n_estimators.' Moreover. The second base learner, Logistic regression: "C," np. Logspace, "penalty":12. To address the problem of an imbalanced data set, which causes overfitting and inconsistent results, we applied the Extreme Gradient Boosting model as a meta-learner, which is counted on an ensemble learning method that allows us to deal with unbalanced dataset classes. A cross-validation technique was implemented for the proposed stacking model using default 5-k folds; they were also included through the GridSearchCV of hyperparameters for the base learners and the meta learner. Finally, the proposed stacking model was verified on a new dataset containing 100,000 records.

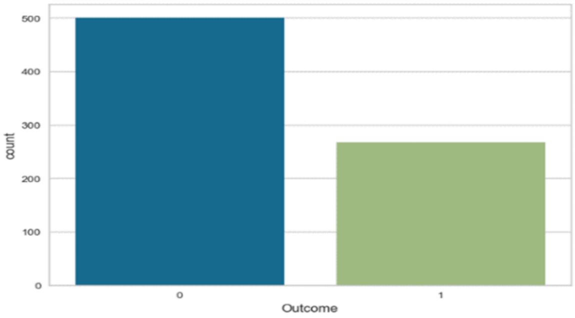


Fig. 4. Pima dataset outcome targets.

TABLE III. STATISTICAL DISTRIBUTION OF PIMA DATASET FEATURES

	Pregnancies	**Glucose**	**Blood Pressure**	**Skin Thickness**	**Insulin**	**BMI**	**Diabetes Pedigree Function**	**Age**	**Outcome**
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.348958
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

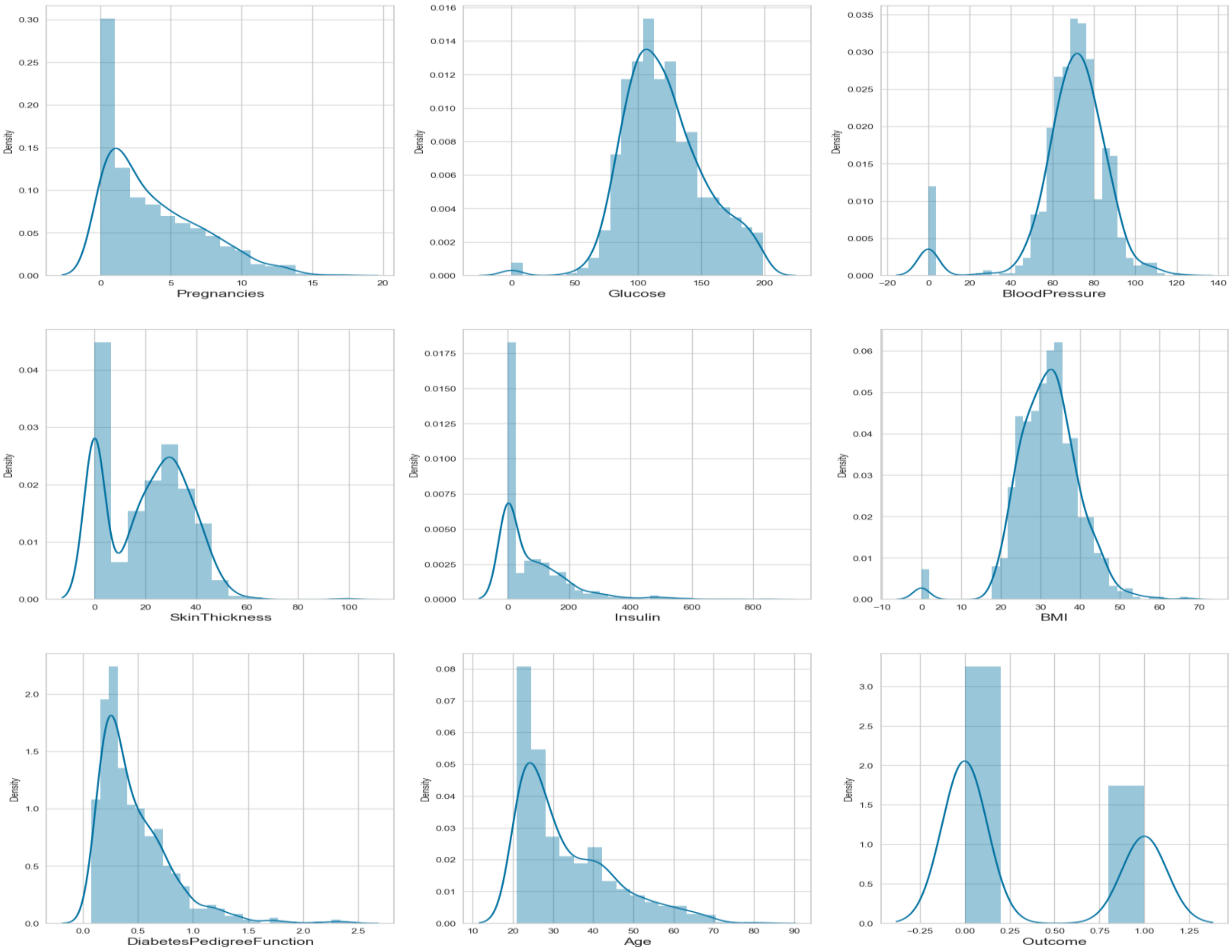
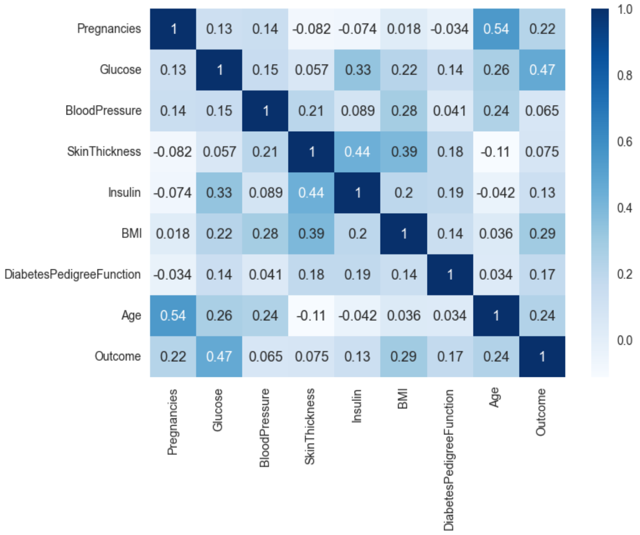


Fig. 5. Pima Dataset features chart .



	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
Pregnancies	1	0.13	0.14	-0.082	-0.074	0.018	-0.034	0.54	0.22
Glucose	0.13	1	0.15	0.057	0.33	0.22	0.14	0.26	0.47
BloodPressure	0.14	0.15	1	0.21	0.089	0.28	0.041	0.24	0.065
SkinThickness	-0.082	0.057	0.21	1	0.44	0.39	0.18	-0.11	0.075
Insulin	-0.074	0.33	0.089	0.44	1	0.2	0.19	-0.042	0.13
BMI	0.018	0.22	0.28	0.39	0.2	1	0.14	0.036	0.29
DiabetesPedigreeFunction	-0.034	0.14	0.041	0.18	0.19	0.14	1	0.034	0.17
Age	0.54	0.26	0.24	-0.11	-0.042	0.036	0.034	1	0.24
Outcome	0.22	0.47	0.065	0.075	0.13	0.29	0.17	0.24	1

Fig. 6. Pima dataset features correlation heatmap.

V. PERFORMANCE MEASURE

To assess the performance of classification models in machine learning and data analysis, we utilize the following metrics:

Accuracy: This metric represents the ratio of correctly predicted samples to the total number of samples. It measures the model's ability to accurately classify both positive and negative cases. However, it's important to note that accuracy can be misleading when dealing with imbalanced classes, as high accuracy can be achieved without focusing on positive classification [13].

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

The output is either diabetic (+dm) or not diabetic (-dm).

- True positive (TP): Prediction is +dm and X is diabetic.
- True negative (TN): Prediction is -dm and X is not diabetic.

- False positive (FP): Prediction is +dm and X is not diabetic.
- False negative (FN): Prediction is -dm and X is diabetic.

Precision: Precision measures the accuracy of predicting positive cases. A high precision value indicates that the model correctly classifies cases as positive when it claims they are [13].

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

The output is either diabetic (+dm) or not diabetic (-dm)

- True positive (TP): Prediction is +dm and X is diabetic.
- False positive (FP): Prediction is +dm and X is not diabetic.

Recall: Also known as sensitivity or true positive rate, recall measures the model's ability to identify all available positive cases. A high recall value signifies that the model can identify most positive cases [13].

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

The output is either diabetic (+dm) or not diabetic (-dm)

- True positive (TP): Prediction is +dm and X is diabetic.
- False negative (FN): Prediction is -dm and X is diabetic.

Cohen's Kappa Score: This metric measures the agreement between two raters. In the context of evaluating classification models, Cohen's Kappa score gauges the agreement between the model's classification and the actual classification. It proves particularly useful when dealing with imbalanced classes or when the model randomly selects between classes [13].

$$CKS = \frac{P_0 - P_e}{1 - P_e} \tag{4}$$

TABLE IV. THE STACKING MODEL RESULTS

	Model	Score
0	Random Forest	0.750730
1	Logistic Regression	0.773706
2	Stacking Model	0.828571

TABLE V. BASE AND META LEARNERS RESULTS

Targets	Precision	Recall	F1-score support	support
0	0.80	0.95	0. 87	41
1	0.90	0. 66	0. 76	29
Accuracy			0.83	70
Macro avg	0.85	0.80	0.81	70
Weighted avg	0.84	0.83	0.82	70

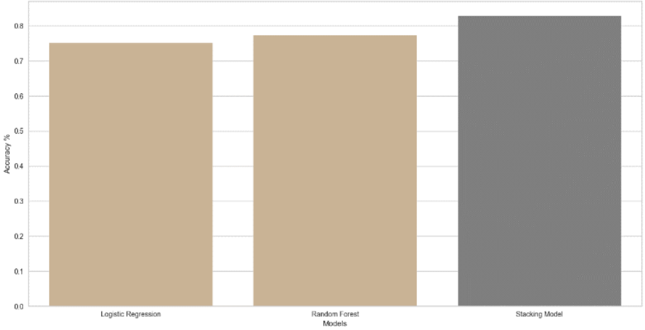


Fig. 7. Base and meta learners results.

where, P_0 represents the accuracy of the models and P_e denotes the agreement between the predicted and actual labels [13].

These metrics aid in comprehending the performance of classification models and identifying their strengths and weaknesses. It is recommended to employ a variety of these metrics to obtain a comprehensive understanding of the model's performance.

A. Validation Dataset

We validated our proposed stacking model performance on a new (binary classification outcomes) diabetes dataset. Diabetes prediction dataset (DPD) is a public dataset consisting of electronic health records (EHRs) that contain digital copies of health records for patients' medical history, diagnosis, therapy, and outcomes. EHRs data is gathered and kept by healthcare providers such as medical centers and hospitals as part of their usual clinical practice. DPD has approximately 100,000 patient records, contributing to significantly measuring the proposed model performance. Its available on Kaggle worldwide datasets repository, link: https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset. DPD dataset features are shown in Table VI; moreover, the correlation heatmap between features is displayed in Fig. 8, whereas Table VII shows the results.

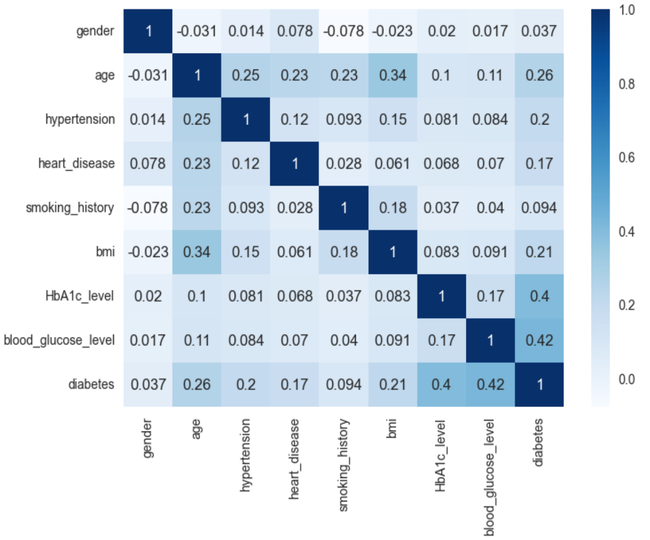


Fig. 8. DPD Dataset features correlation heatmap.

TABLE VI. DPD DATASET INFORMATION

Data columns (total 9 columns):		
Column	Non-Null Count	Datatype
Gender	100,000 non-null	object
Age	100,000 non-null	float64
hypertension	100,000 non-null	int64
Heart disease	100,000 non-null	int64
Smoking history	100,000 non-null	object
BMI	100,000 non-null	float64
HbA1c_level	100,000 non-null	float64
Blood_glucose_level	100,000 non-null	int64
Outcomes	768 non-null	int64

TABLE VII. THE STACKING MODEL RESULTS ON DPD DATASET

	Model	Score
0	Random Forest	0.95064
1	Logistic Regression	0.97012
2	Stacking Model	0.97160

VI. RESULTS

In this research, we built a stacking ensemble model to predict diabetes mellitus using a combination of machine learning models; where random forest, logistic regression models were applied as base learners and the extreme gradient Boosting model as meta learner, techniques such as cross validation and GridSearchCV were applied. We also replaced the zeros in the Pima dataset with values (median - mean) according to the types of data distribution with features columns (normal - skewed), as mentioned in [71] that if we remove zero values, the performance will improve. We obtained an accuracy of 83% in predicting diabetes mellitus with Pima dataset, and we also verified the efficiency of the proposed model on a large dataset containing approximately 100,000 records, with accuracy of 97%, where kapa Cohen score was 61% on Pima dataset, and 78% on DPD dataset. More details are discussed in the next paragraph. We observe that our proposed ensemble stacking model for predicting diabetes covers the shortcomings mentioned in Table I, such as the study of S. Härner and D. Ekman (2022) regarding the need for using hyperparameters search optimizer to improve model results. Moreover, a study of H. Syed and T. Khan (2020) about the geographical scope of the study dataset, where we used two different dataset scopes. The Table IV shows the detailed results of the proposed stacking model. In addition, Table V, Fig. 7 shows the base and meta learners results.

VII. DISCUSSION

A. *Results with XGBoost and GridSearchCV*

In this experiment, we utilized a combination of ML and DL classifiers to predict diabetes mellitus. Fig. 1 illustrates the stacking model methodology in this experiment, where we initiated each of the RF, LR models as base learners and the XGB classifier as a Meta-learner. At the same time, we used GridSearchCV Hyperparameters optimizer to find the optimal results for the Random Forest classifier using the following hyperparameters: bootstrap, max_features, min_samples_leaf, n_samples, split,n_estimators Moreover. The second base learner, Logistic regression: "C," np. Logspace, "penalty":12. To address the problem of an imbalanced data set, which causes overfitting and inconsistent results, we applied the Extreme Gradient Boosting model as a meta-learner, which is counted on an ensemble learning method that allows us to deal with unbalanced dataset classes. A cross-validation technique was implemented for the proposed stacking model using default 5-k folds; they were also included through the GridSearchCV of hyperparameters for the base learners and the meta learner. The results were as follows: The prediction accuracy of the stacking ensemble model on Pima dataset is 83%, kapa Cohen score 61%, where on DPD dataset was 97% accuracy and 78% kapa Cohen score. Table IV shows the results in detail. Fig. 9 displays the differences in results between Pima and DPD datasets.

B. *Comparative Analysis with Existing Work*

1) First study: S. Härner and D. Ekman (2022) [34] proposed an ensemble stacking model for predicting diabetes using a combination of machine learning models, including (Decision Tree and Naive Bayes models). The Pima dataset was used in this study, and the results indicated that the proposed stacking model can predict diabetes with 75.56% accuracy. In addition, it was mentioned that there were limitations during the study, such as not using an optimizer to search in hyperparameters to find the best results for base learners in the stacking model.

2) Second study: Patil et al. (2023) [20] Suggested an ensemble stacking model for predicting diabetes, using a combination of machine learning models such as (decision tree, naïve Bayes (NB), multilayer perceptron (MLP), SVM, and KNN). The Pima dataset was also used in this study. The results indicated that the stacking model can predict diabetes with 81.9% accuracy. In addition, we noticed that they never mentioned the cross-validation technique during the proposed methodology, which plays an essential role in building the stacking model. Moreover, no optimizer was used in searching the hyperparameters while training the base learners' models to get better results.

3) Third study: Lei Qin (2022) [35] devised an ensemble stacking approach to predict diabetes. They amalgamated various machine learning models—Logistic Regression, K-Nearest Neighbors, Decision Trees, Gaussian Naive Bayes, and Support Vector Machine (SVM). Utilizing the Pima dataset, their findings revealed that the stacking model achieved an 81.63% accuracy in diabetes prediction. However, the absence of an optimizer for hyperparameter tuning during base learner model training might have hindered the quest for better outcomes. Additionally, the limited size of the dataset posed a challenge, potentially impacting the attainment of optimal results.

4) Forth study: Kumari et al. (2021) [36] suggested an ensemble soft voting model for predicting diabetes, using a combination of machine learning models such as (Random

Forest (RF), logistic regression (LR), and Naive Bayes (NB)). The Pima dataset was also used in this study. The results indicated that the soft voting model can predict diabetes with 79.04% accuracy. Furthermore, it's important to highlight that the proposed methodology overlooked the inclusion of cross-validation, a crucial technique integral to ensuring robustness by assessing the performance of individual models across various subsets of the data, thereby refining their predictions' collective contribution to the ensemble. Additionally, the absence of an optimizer in the pursuit of hyperparameter tuning during the training of base learner models might have impacted the potential for achieving superior results.

We observe that our proposed ensemble stacking model outperforms in predicting diabetes accuracy compared to other proposed models in the [20], [34], [35], [36] studies, in our approach, we leveraged the GridsearchCV optimizer to search for the best hyperparameters for our base learners. Interestingly, this optimization technique wasn't utilized in either the First Study or the Second Study. This optimization significantly boosted our base learners' learning process, leading to extracting the most optimal results possible. Furthermore, the second and fourth studies overlooked the utilization of cross-validation—a critical technique for evaluating a predictive model's generalization capacity. In contrast, our model applied this method, dividing the dataset into k-folds during both training and testing. This implementation effectively evaluated and prevented overfitting, significantly enhancing our prediction model's performance.

Table VIII meticulously delineates and highlights the disparities and advantages between these studies, emphasizing the significant enhancements our approach brings to the table in comparison to the methodologies adopted in the First Study and the Second Study.

TABLE VIII. COMPARISON WITH EXISTING STUDIES

Authors	Techniques used	Dataset	Accuracy
S. Härner and D. Ekman (2022)	stacking ensemble approach (Decision tree (DT), Naïve Bayes (NB)), Cross-validation	Pima dataset	75.56%
Patil et al (2023)	Stacking ensemble approach (Decision tree (DT), Naïve Bayes (NB), multilayer perceptron (MLP), SVM, and KNN)	Pima dataset	81.9%
Lei Qin (2022)	Stacking ensemble approach (Logistic Regression, K-Nearest Neighbors, Decision Trees, Gaussian Naive Bayes, and Support Vector Machine (SVM))	Pima dataset	81.63%
Kumari et al (2021)	Soft voting ensemble approach (Random Forest (RF), Logistic regression (LR), and Naive Bayes (NB))	Pima dataset	79.04%
Our proposed model	Stacking ensemble approach (Random Forest, Logistic Regression, XGboost) GridSearchCV, Cross-validation.	Pima Dataset	83%
Our proposed model on the validation dataset	Stacking ensemble approach (Random Forest, Logistic Regression, XGboost) GridSearchCV, Cross-validation.	DPD Dataset	97%

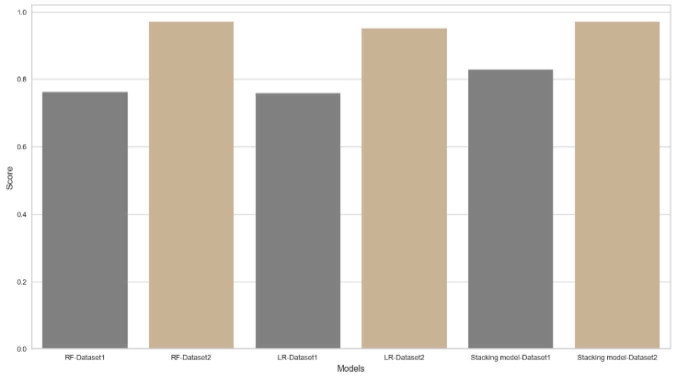


Fig. 9. Comparison results between pima and DPD datasets.

VIII. CONCLUSION

Diabetes mellitus (DM) is a common disease that threatens the health of society, causing many serious diseases such as kidney failure, heart disease, and blindness. In this research, we proposed a novel stacking ensemble model to predict diabetes mellitus using a Pima dataset and combined machine learning models, where we used the Random Forest (RF) and Logistic Regression (LR) as base learners models and XGBoost as a Meta-Learner model. Moreover, we applied the cross-validation technique to get the optimal results in the RF, LR, models through the Grid Search optimizer technique. To avoid the problem of an imbalanced dataset, which causes overfitting and inconsistent results, we applied the XGBoost model as a meta-learner. However, the dataset has been cleaned from zero values that harm the prediction result, which was illogical to have zero values on some columns, like glucose in the blood. To address this problem, we replaced zero values with median and mean values based on the type of distribution (normal - skewed). The results indicate that our proposed stacking model can predict diabetes mellitus with an accuracy of 83% with the Pima dataset, and 97% on the DPD dataset. As recommendations, our stacking model can be applied in a diagnostic application for diabetes mellitus; in addition, it can be tested on a new huge and diverse dataset to obtain more accurate results. Moreover, we can use deep-learning models to generate new patterns that help us diagnose DM robustly, which also can happen with different types of diabetes, such as type 1 and type 2 diabetes and gestational diabetes.

REFERENCES

[1] H. Sone, 'Diabetes Mellitus', in Encyclopedia of Cardiovascular Research and Medicine, R. S. Vasan and D. B. Sawyer, Eds., Oxford: Elsevier, 2018, pp. 9–16. doi: https://doi.org/10.1016/B978-0-12-809657-4.99593-0.

[2] T. Andoh, 'Subchapter 19A - Insulin', in Handbook of Hormones, Y. Takei, H. Ando, and K. Tsutsui, Eds., San Diego: Academic Press, 2016, pp. 157-e19A-3. doi: https://doi.org/10.1016/B978-0-12-801028-0.00148-3.

[3] J. Hippisley-Cox and C. Coupland, 'Diabetes treatments and risk of amputation, blindness, severe kidney failure, hyperglycaemia, and hypoglycaemia: open cohort study in primary care', BMJ, p. i1450, Mar. 2016, doi: 10.1136/bmj.i1450.

[4] A. N. Baanders and M. J. W. M. Heijmans, 'The Impact of Chronic Diseases: The Partner's Perspective', Family & Community Health, vol. 30, no. 4, pp. 305–317, Oct. 2007, doi: 10.1097/01.FCH.0000290543.48576.cf.

[5] P. Saeedi et al., ‘Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition’, Diabetes Research and Clinical Practice, vol. 157, p. 107843, Nov. 2019, doi: 10.1016/j.diabres.2019.107843.

[6] C. V. A. Collares et al., ‘Transcriptome meta-analysis of peripheral lymphomononuclear cells indicates that gestational diabetes is closer to type 1 diabetes than to type 2 diabetes mellitus’, Mol Biol Rep, vol. 40, no. 9, pp. 5351–5358, Sep. 2013, doi: 10.1007/s11033-013-2635-y.

[7] A. E. Butler and D. Misselbrook, ‘Distinguishing between type 1 and type 2 diabetes’, BMJ, p. m2998, Aug. 2020, doi: 10.1136/bmj.m2998.

[8] World Health Organization, Global report on diabetes. Geneva: World Health Organization, 2016. Accessed: Jan. 25, 2023. [Online]. Available: https://apps.who.int/iris/handle/10665/204871

[9] A. Thammano and A. Meengen, ‘A New Evolutionary Neural Network Classifier’, in Advances in Knowledge Discovery and Data Mining, T. B. Ho, D. Cheung, and H. Liu, Eds., in Lecture Notes in Computer Science, vol. 3518. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 249–255. doi: 10.1007/11430919_31.

[10] Y. LeCun, Y. Bengio, and G. Hinton, ‘Deep learning’, Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[11] T. G. Dietterich, ‘Ensemble Methods in Machine Learning’, in Multiple Classifier Systems, in Lecture Notes in Computer Science, vol. 1857. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15. doi: 10.1007/3-540-45014-9_1.

[12] L. I. Kuncheva, Combining pattern classifiers: methods and algorithms, Second edition. Hoboken, NJ: Wiley, 2014.

[13] M. Gollapalli et al., ‘A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM’, Computers in Biology and Medicine, vol. 147, p. 105757, 2022, doi: https://doi.org/10.1016/j.compbiomed.2022.105757.

[14] A. Dutta et al., ‘Early Prediction of Diabetes Using an Ensemble of Machine Learning Models’, IJERPH, vol. 19, no. 19, p. 12378, Sep. 2022, doi: 10.3390/ijerph191912378.

[15] S. M. Ganie and M. B. Malik, ‘An ensemble Machine Learning approach for predicting Type-II diabetes mellitus based on lifestyle indicators’, Healthcare Analytics, vol. 2, p. 100092, Nov. 2022, doi: 10.1016/j.health.2022.100092.

[16] U. e Laila, K. Mahboob, A. W. Khan, F. Khan, and W. Taekeun, ‘An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning: An Empirical Study’, Sensors, vol. 22, no. 14, p. 5247, Jul. 2022, doi: 10.3390/s22145247.

[17] D. Pankaj Javale and S. Suhas Desai, ‘Machine learning ensemble approach for healthcare data analytics’, IJEECS, vol. 28, no. 2, p. 926, Nov. 2022, doi: 10.11591/ijeecs.v28.i2.pp926-933.

[18] A. Singh, A. Dhillon, N. Kumar, M. S. Hossain, G. Muhammad, and M. Kumar, ‘eDiaPredict: An Ensemble-based Framework for Diabetes Prediction’, ACM Trans. Multimedia Comput. Commun. Appl., vol. 17, no. 2s, pp. 1–26, Jun. 2021, doi: 10.1145/3415155.

[19] G. Geetha and K. M. Prasad, ‘An Hybrid Ensemble Machine Learning Approach to Predict Type 2 Diabetes Mellitus’, WEB, vol. 18, no. Special Issue 02, pp. 311–331, Apr. 2021, doi: 10.14704/WEB/V18SI02/WEB18074.

[20] R. N. Patil, S. Rawandale, N. Rawandale, U. Rawandale, and S. Patil, ‘An efficient stacking based NSGA-II approach for predicting type 2 diabetes’, IJECE, vol. 13, no. 1, p. 1015, Feb. 2023, doi: 10.11591/ijece.v13i1.pp1015-1023.

[21] A. H. Syed and T. Khan, ‘Machine Learning-Based Application for Predicting Risk of Type 2 Diabetes Mellitus (T2DM) in Saudi Arabia: A Retrospective Cross-Sectional Study’, IEEE Access, vol. 8, pp. 199539–199561, 2020, doi: 10.1109/ACCESS.2020.3035026.

[22] D. Berrar, ‘Cross-Validation’, in Encyclopedia of Bioinformatics and Computational Biology, Elsevier, 2019, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.

[23] M. Maalouf, ‘Logistic regression in data analysis: an overview’, IJDATS, vol. 3, no. 3, p. 281, 2011, doi: 10.1504/IJDATS.2011.041335.

[24] T. Hastie, 'The Elements of Statistical Learning: Data Mining, Inference, and Prediction'. 01 2009. doi: 10.1007/978-0-387-84858-7.

[25] R. Tibshirani, ‘Regression Shrinkage and Selection via the Lasso’, Journal of the Royal Statistical Society. Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996.

[26] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[27] C.-C. Chang, Y.-Z. Li, H.-C. Wu, and M.-H. Tseng, ‘Melanoma Detection Using XGB Classifier Combined with Feature Extraction and K-Means SMOTE Techniques’, Diagnostics, vol. 12, no. 7, p. 1747, Jul. 2022, doi: 10.3390/diagnostics12071747.

[28] Z. Zhao, H. Peng, C. Lan, Y. Zheng, L. Fang, and J. Li, ‘Imbalance learning for the prediction of N6-Methylation sites in mRNAs’, *BMC Genomics*, vol. 19, no. 1, p. 574, Dec. 2018, doi: 10.1186/s12864-018-4928-y.

[29] N. H. N. B. M. Shahri, S. B. S. Lai, M. B. Mohamad, H. A. B. A. Rahman, and A. B. Rambli, ‘Comparing the Performance of AdaBoost, XGBoost, and Logistic Regression for Imbalanced Data’, *ms*, vol. 9, no. 3, pp. 379–385, May 2021, doi: 10.13189/ms.2021.090320.

[30] F. Pedregosa et al., ‘Scikit-learn: Machine Learning in Python’, Journal of Machine Learning Research, vol. 12, no. 85, pp. 2825–2830, 2011

[31] R. D. Joshi and C. K. Dhakal, ‘Predicting Type 2 Diabetes Using Logistic Regression and Machine Learning Approaches’, *IJERPH*, vol. 18, no. 14, p. 7346, Jul. 2021, doi: 10.3390/ijerph18147346.

[32] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, ‘Predicting Diabetes Mellitus With Machine Learning Techniques’, *Front. Genet.*, vol. 9, p. 515, Nov. 2018, doi: 10.3389/fgene.2018.00515.

[33] L. Breiman, ‘Random Forests’, Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[34] S. Härner and D. Ekman, Comparing Ensemble Methods with Individual Classifiers in Machine Learning for Diabetes Detection. 2022.

[35] L. Qin, ‘A Prediction Model of Diabetes Based on Ensemble Learning’, in *Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition*, Xiamen China: ACM, Sep. 2022, pp. 45–51. doi: 10.1145/3573942.3573949.

[36] S. Kumari, D. Kumar, and M. Mittal, ‘An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier’, *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40–46, Jun. 2021, doi: 10.1016/j.ijcce.2021.01.001.

Encryption Traffic Classification Method Based on ConvNeXt and Bilinear Attention Mechanism

Xiaohua Feng[1], Yuan Liu[2]
School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China[1]
Jiangsu Key Laboratory of Media Design and Software Technology, Wuxi, China[2]

***Abstract*—The rapid growth in internet traffic resulted to the emergence of network traffic categorization as a crucial area of research in network performance and management. This technological advancement has demonstrated its efficacy in aiding network administrators to identify anomalies within network behavior. However, the widespread adoption of encryption technology and the continual evolution of encryption protocols present a novel challenge in the classification of encrypted traffic. Addressing this challenge, this paper introduces an innovative methodology for classifying encrypted traffic by harnessing ConvNeXt and a fusion attention mechanism. Through the representation of traffic data as images and the integration of a bilinear attention mechanism into the model, our proposed approach attains heightened precision in the classification of encrypted network traffic. To substantiate the effectiveness of our methodology, experiments were conducted employing the publicly available ISCX VPN-nonVPN dataset. The experimental findings showcase superior recognition performance, underscoring the efficacy of the proposed approach.**

Keywords—Encryption traffic recognition; end-to-end; convolutional neural network; bilinear attention module

I. Introduction

As the number of Internet-connected devices continues to rise, the overall volume of network traffic is experiencing a significant expansion. Consequently, the significance of network traffic classification in contemporary network management has become increasingly evident. Moreover, with the prevalence of encrypted traffic in modern network applications, the task of classifying encrypted network traffic has assumed a pivotal role in network management. This technology has proven to be effective in assisting administrators in identifying and locating network anomalies, detecting network security threats, and is widely applied in specific areas such as network management, security monitoring, and Quality of Service (QoS) management.

In previous studies, researchers have proposed various methods to tackle the challenge of encrypted network traffic classification. These methods encompass a range of approaches, including port-based methods [1], payload-based traffic methods [2], and machine learning-based approaches [3]. These approaches have made notable contributions to the field, but they also possess certain limitations and drawbacks that warrant further exploration and improvement.

The port classification method, which is based on the transport layer, primarily categorizes ports according to the standards provided by the Internet Assigned Numbers Authority (IANA). For instance, ports like 80 and 443 are commonly utilized as defaults in web services. Port-based solutions, although simple and rapid, have proven insufficient in the present context due to the rise of network protocols and technologies like dynamic ports.

On the other hand, payload-based methods analyze traffic by examining keywords or patterns in data packets. However, the use of encryption technology for network traffic, driven by concerns for security and privacy, renders this method less applicable. In traditional machine learning methods, the performance of algorithms heavily relies on features designed by professionals for different types of traffic. However, in the complex traffic landscape of today, these designed traffic features are unable to cover all categories of traffic.

In recent years, with the significant development of deep learning in the field of image processing, many researchers have also applied deep learning to traffic classification. Deep learning reduces the need for manually designed features by directly learning features from the complex traffic patterns. This approach has shown promise in improving the accuracy and efficiency of traffic classification.

Nevertheless, with the rapid growth of the Internet and the increasing complexity of network applications, traditional traffic classification methods are facing more and more challenges. There is a need for innovative approaches and methodologies that can effectively handle the evolving nature of network traffic. Integrating interdisciplinary methods and exploring new theoretical frameworks may offer potential solutions to overcome these challenges and enhance the accuracy and efficiency of traffic classification in the future.

Deep convolutional neural networks (CNNs) have made remarkable progress in computer vision and natural language processing due to the advancements in deep learning. In the domain of network traffic classification, convolutional neural networks are also widely applied. A notable architecture ConvNeXt [4], has demonstrated comparable performance to models like ResNet while maintaining a relatively low number of parameters. Additionally, the self-attention mechanism, proposed by Google, has emerged as an alternative to traditional recurrent neural networks (RNNs), offering lower computational complexity and enhanced focus on crucial features.

The paper presents a novel model that integrates ConvNeXt with attention processes to improve the precision of network traffic categorization. This paper introduces many novel advancements:

- A bilinear attention mechanism module is proposed that effectively enhances the accuracy of convolutional neural networks in fine-grained tasks. This module enables the model to concentrate on important attributes and enhances the overall performance.
- A novel method for encrypting traffic classification based on the ConvNeXt fusion attention mechanism. By leveraging information from different scales in the network, ConvNeXt improves model performance. The introduction of the attention mechanism further enhances the model's ability to accurately learn key information in network traffic, prioritizing features that are crucial for classification.
- Adapt the new ConvNeXt model's parameters to traffic classification methods and conducting experiments on multiple datasets. Through these experiments, the feasibility and effectiveness of the proposed method in traffic classification tasks are demonstrated.

This paper is organized into four Sections. Section II provides a review of prior work and introduces the ConvNeXt network. Section III presents the proposed methodology, including the bidirectional linear attention mechanism module. Section IV focuses on the experimental aspects, describing the setup, dataset, and analysis of results. Finally, this work is concluded in the Section VI where possible future research directions are indicated.

II. Related Work

Previous studies have proposed numerous mature methods for network traffic classification. This section will elaborate on these classification methods, providing an overview of the approaches developed by researchers. It aims to provide a concise yet comprehensive understanding of the existing techniques used in this field.

A. The Methods Based on Payload and Machine Learning Approaches

With the emergence of new network protocols and dynamic ports, traditional port-based classification methods have become inadequate for the current network environment. In response, researchers have proposed Deep Packet Inspection (DPI) technology. DPI technology goes beyond traditional packet inspection of the first four layers of IP packets and reads and reassembles the application layer data to achieve traffic classification. P. Khandait et al. [5] introduced a system called Length-Based Matching (LBM) which includes an innovative acceleration approach for RegEx matching. Wang et al. [6] introduced a framework called lightweight Deep Packet Inspection (LW-DPI). S. Fernandes, R. Antonello et al. [7] introduced a Bitcoding system, which is a method for generating traffic classification signatures at the bit level using DPI. These paper findings have shown commendable performance. However, DPI technology faces two inherent challenges.

The first challenge is privacy concerns. In the current era of the Internet, users have become increasingly concerned about the privacy of their transmitted data. DPI technology involves reading the content of user transmissions, which inevitably violates user privacy. This raises ethical and legal concerns, as users expect their data to remain confidential and secure.

The second challenge is encryption. In response to the growing apprehension among users regarding the confidentiality of network traffic, a significant number of products have adopted the practice of encrypting their network traffic. According to Google's Transparency Report, over 90% of traffic in Google's products is encrypted. Encrypted traffic does not allow its data to be read, rendering DPI detection methods ineffective for classification. As more and more internet traffic becomes encrypted for security reasons, the limitations of DPI technology become increasingly apparent.

In light of these challenges, machine learning has emerged as a promising solution to the traffic classification problem. In stark contrast to DPI, machine learning-based methods classify traffic by learning statistical features from the traffic data. Classification schemes in classical machine learning are classified according to the degree of supervision, which includes supervised, unsupervised, and semi-supervised methods. Using supervised machine learning, K.L. Dias [8] explored real-time video traffic to categorize real-time applications. Feature engineering and classifier construction were employed by Cao et al. [9] in order to increase the accuracy of traffic categorization in the SVM-based model. Using K-means clustering as a tool for unsupervised learning, Wang et al. [10] examined the effectiveness of grouping network flows based on similarity using the algorithm. Höchst et al. [11] developed a method for classifying traffic flow by utilizing statistical characteristics derived from a neural autoencoder algorithm. Using semi-supervised learning, B. Ghita et al. [12] used two-phase learning approach for traffic classification. The semi-supervised method developed by F. Noorbehbahani et al. [13] consists of Clustering using X-means and propagation of labels. However, the effectiveness of machine learning-based classifiers heavily relies on well-designed features. Designing optimal features requires experienced professionals to invest a significant amount of time in manual design. Furthermore, the overall generalization performance of such methods is relatively poor. Additionally, due to the involvement of human intervention, these methods require professionals to adapt the classifier when the current environment changes or when it is not suitable for a different distribution of similar datasets. This reliance on human intervention can be a cumbersome and time-consuming process.

In summary, while DPI technology offers a more in-depth approach to traffic classification, it faces challenges related to privacy concerns and the rise of encryption. Machine learning-based methods, on the other hand, provide a promising alternative, but they require optimal feature design and can be limited by the need for human intervention.

B. The Methods Based on Deep Learning

In contrast to machine learning methods, deep learning methods eliminate the reliance on manually designed features. Deep learning operates through an end-to-end process, where raw data is directly inputted into the model. The model then autonomously learns its own features based on the outcomes, facilitating global optimization. Although deep learning necessitates a substantial amount of training data, it plays a pivotal role in reducing human intervention throughout the entire workflow. Employing specialized techniques for raw data processing can effectively enhance performance. Shapira, Tal [14] approached the problem by treating packet size and arrival time in network traffic as correlation graphs. They eliminated packets larger than 1500 bytes, focusing on packets arriving within the first 60 seconds. This mapping generated a 1500x1500 histogram, which was then fed into a convolutional neural network based on the LeNet-5 architecture for classification. Lan, Jinghong [15] and others tackled the issue from multiple feature levels, employing different extraction methods for each feature and subsequently processing the combined features. D'Angelo, Gianni [16] proposed a method of data collection through window sampling, followed by statistical analysis of its features. This approach primarily involved counting various data exchanged within the window, such as the number and length of packets. The information was then inputted into an SAE network for classification.

Achieving satisfactory results can also be accomplished by combining multiple models to identify and classify traffic features. Wang, Wei [17] [18] transformed traffic into images and employed one-dimensional convolutional neural networks and representation learning. Both methods demonstrated strong performance. Maonan, Wang [19] combined ResNet and AutoEncoder models for classification. They divided the data, extracting the original data into images as input for feature extraction in the ResNet network. They also inputted the statistical features of the data flow into a network based on AutoEncoder to extract statistical features. The two sets of features were then combined into complex features for classification.

Additionally, attention mechanisms have proven effective in enhancing model performance. Attention mechanisms can be integrated with traditional convolutional neural networks or recurrent neural networks to augment the capabilities of the original models. Barut, Onur [20] incorporated position and time information for each bit into the original PCAP stream data. They utilized a model that combines multi-head self-attention pooling and 2D CNN for classification. Liu, Xun [21] utilized intercepted packets as input data and employed a model that combines attention and GRU for classifying network traffic. Xiao, Xi [22] utilized side channel data to improve algorithm performance. They inputted this data into a model consisting of RNN and attention mechanisms for classification.

C. ConvNeXt Convolutional Neural Network

With the continuous advancements in deep learning, the field of image classification has witnessed the emergence of new algorithm networks. Among these networks, Swin Transformer has been at the forefront of fine-grained classification since 2020, demonstrating exceptional performance and gradually replacing the conventional convolutional neural networks. However, in 2022, scholars such as Liu and Zhuang [4] proposed ConvNeXt, a novel approach that builds upon the Swin Transformer by studying its layer structure, downsampling methods, activation functions, and data processing techniques. This research has led to significant improvements in the accuracy of convolutional neural networks, reaffirming their crucial role in image classification.

The network architecture of ConvNeXt is not notably distinctive, as it integrates various components that have been previously employed in research. Fig. 1 exhibits similarities to the layer arrangement of ResNet50. ConvNeXt incorporates downsampling techniques and layer normalization, while also sharing commonalities in its recurrence approach and structural combination with ResNet50. Fig. 2 can be juxtaposed with the layer architecture of MobileNetV2. Inverted residuals and the Swin Transformer's MLP structure are combined to form ConvNeXt blocks. ConvNeXt improves the accuracy of coarse-grained image categorization by combining the processing methods of the Swin Transformer with the inherent properties of convolution.

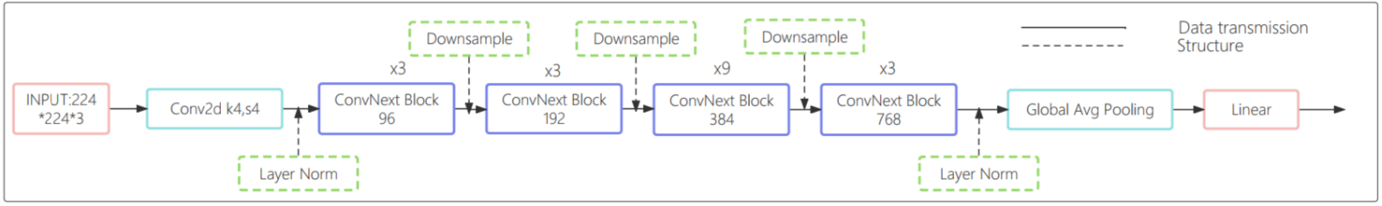


Fig. 1. Network diagram chart of ConvNeXt.

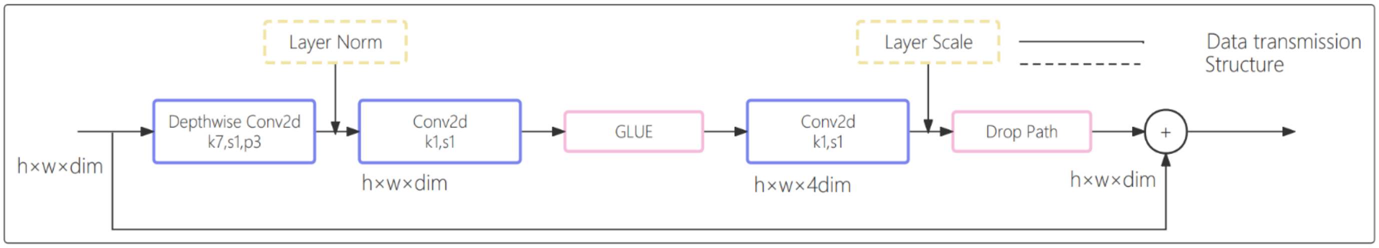


Fig. 2. Structure diagram of ConvNeXt block.

III. Methods

This section provides an overview of the dataset utilized in this paper, the data processing methodology, the construction of the model, and the parameter configuration within the model.

A. Data Preprocessing

The dataset utilized in this work consists of capture files, with each file corresponding to a distinct program, traffic type, and encryption method. These capture files are stored in the PCAP format. The initial 24 bits of a PCAP packet contain crucial data information, including a 4-byte file magic number, a 2-byte major version number, a 2-byte minor version number, a 4-byte local timezone, a 4-byte timestamp, a 4-byte maximum storage, and a 4-byte link type. Subsequently, each file consists of a combination of packet headers and packet data. The packet header is composed of four 4-byte fields, namely the high-order timestamp, low-order timestamp, current packet length, and offline data length. The data structure is visually depicted in Fig. 3.

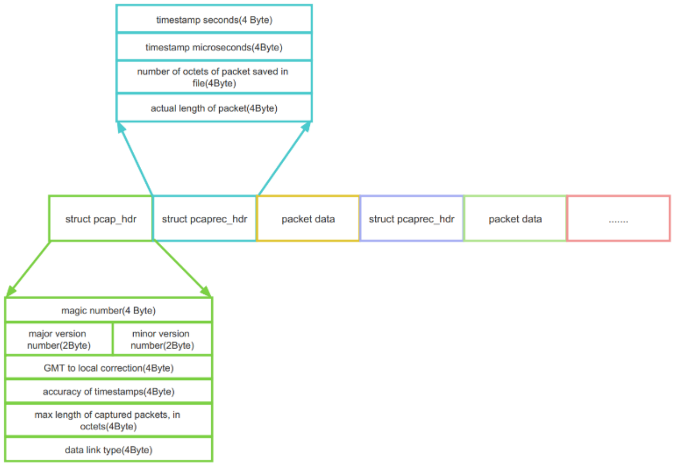

timestamp seconds(4 Byte)
timestamp microseconds(4Byte)
number of octets of packet saved in file(4Byte)
actual length of packet(4Byte)
struct pcap_hdr
struct pcaprec_hdr
packet data
struct pcaprec_hdr
packet data
....
magic number(4 Byte)
major version number(2Byte)
minor version number(2Byte)
GMT to local correction(4Byte)
accuracy of timestamps(4Byte)
max length of captured packets, in octets(4Byte)
data link type(4Byte)

Fig. 3. Structure of pcap file.

In the data cleaning section, this paper undertakes the removal of IP addresses and port information from the source data. This step is crucial to prevent the model from relying on these details and making incorrect judgments, thereby eliminating biases associated with IP addresses and ports. Additionally, as the data remains in the PCAP format even after being divided into flows, the PCAP header also needs to be eliminated during the data cleaning process.

This work utilizes the five-tuple {source IP, source port, destination IP, destination port, protocol} to divide each PCAP file into numerous unidirectional flows. Subsequently, the data undergoes transformation into images through the following key steps:

1) Data refinement: Duplicate or blank data can significantly impact the training of deep learning models, introducing biases in learning features and reducing classification accuracy. Hence, it is imperative to remove any duplicate or blank content present in certain packets.

2) Privacy processing: In order to diminish the model's reliance on IP addresses within the data, this paper employs zero-padding on the IP-related information by filling them with zeros.

3) Data trimming: Based on data analysis, to comprehensively capture feature information of network traffic and achieve more accurate network traffic classification, this paper trims the payload data to a fixed length of 576 bytes. If the file size exceeds 576 bytes, the excess portion is deleted. Conversely, if the length is less than 576 bytes, zeros are added at the end.

4) Data transformation: Each data segment with a length of 576 bytes is transformed into a 24x24 image to facilitate processing by the model.

The aforementioned steps ensure the cleanliness, privacy protection, and appropriate formatting of the data, enabling effective analysis and classification of network traffic. By applying the above processing methods, 24x24 images of each category will be obtained, and some of the images are shown in Fig. 4.

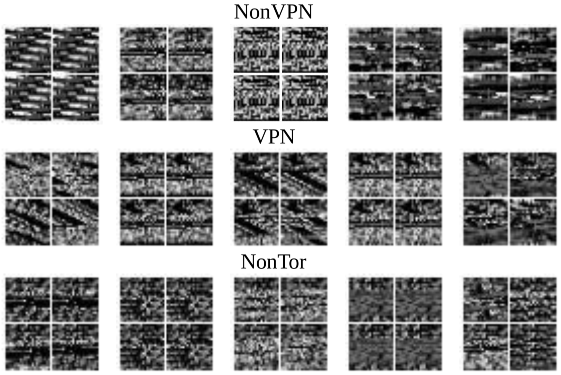


Fig. 4. Partial preprocessed images.

B. Model Structure

The primary model utilized in this paper is ConvNeXt-Tiny, selected for its lower parameter count and enhanced feature extraction capabilities, thereby reducing hardware requirements during training. In the task of traffic classification, the information content of the split pcap files is limited. To maximize the inclusion of information, the length of the traffic is truncated to 576 based on an analysis of the length feature of the original data flow. Subsequently, it is transformed into a 24x24 image.

1) ConvNeXt model incorporating attention mechanism: The network structure of ConvNeXt comprises four different kinds of blocks, each with a different number of channels. Each block type undergoes a specific number of cycles: the first kind goes through three times, resulting in a feature map with 96 dimensions, the second kind goes through three times, resulting in a feature map with 192 dimensions; the third kind goes through three times, resulting in a feature map with 384 dimensions, and the fourth kind goes through three times, resulting in a feature map with 192 dimensions768. Before each block, the feature map is subjected to a downsampling procedure, which decreases the final output size (W, H) by half compared to the original size, while simultaneously increasing the output dimension by convolution. Adding weight values to channels is crucial in this scenario.

Precise localization of the target is crucial in tasks that involve fine-grained categorization, as it enables successful feature extraction. In order to enhance the network's capacity to learn particular target features during training, we provide two methods called Embedded Block Attention (EBA) and Sequential Block Attention (SBA), both of which rely on embedded locations. The ConvNeXt-Tiny network incorporates the attention mechanism into each block, enabling the attention mechanism to be integrated into the internal loop of each block type. The term used to describe this integration is EBA. The feature maps of inverted residuals are subjected to attention weights, and attention parameters within each block are periodically trained. This process results in the formation of a feature map with attention weights for each block's cyclic object. This method allows for the iterative training of attention settings. To obtain a comprehensive depiction of the altered structure of the ConvNeXt Block, refer to Fig. 5.

$$\boldsymbol{F_{a1}} = \sigma(\mathrm{Con\,v}(\mathrm{GeL\,U}(\mathrm{L\,N}(\mathrm{Con\,v}(\mathrm{Conca\,t}(\mathrm{Avgpoo\,l}(\boldsymbol{F}), \mathrm{Maxpoo\,l}(\boldsymbol{F}))))))) = \sigma(\boldsymbol{W_1}(\mathrm{GeLU}(\mathrm{LN}(\boldsymbol{W_0}(\mathrm{Concat}(\boldsymbol{F^c_{avg}}, \boldsymbol{F^c}_{\max})))))) \quad (1)$$

In contrast to EBA, the integration of the attention mechanism through SBA in ConvNeXt-Tiny does not disrupt the cyclic process of the blocks. Instead, it applies the attention mechanism to the feature maps after the cyclic process of each type of block. This allows the training of attention weights to take into account the entire cyclic feature map, rather than individual blocks. As a result, the number of training iterations required for attention parameters is reduced. Additionally, to address challenges such as degradation, gradient explosion, and gradient vanishing in deep networks, residual shortcuts are employed to add the downsampled feature maps to the output of the attention module based on channels. For a more detailed illustration of the specific structures, (see Fig. 6).

When incorporating the attention mechanism using SBA, it is crucial to observe that the remaining connections of attention, except for the third kind of block, are removed (see Fig. 6). The rationale behind this decision stems from the observation that the third kind of block present in all iterations of ConvNeXt experiences the greatest extensive amount of iterations, specifically nine cycles in the Tiny version. Introducing external residual connections would result in the inclusion of a significant amount of untreated and superficial semantic information into the attention feature map. The integration would ultimately diminish the network's overall ability to extract features and learn.

2) Bilinear attention mechanism: Convolutional Block Attention Module (CBAM)[23] is enhanced in this paper, and a new attention mechanism is proposed to improve the accuracy of ConvNeXt. The attention mechanism also allocates feature weights in ConvNeXt-Tiny. This section offers a comprehensive overview of the planned CBAM enhancement.

This paper proposes a Bilinear CBAM (BLCBAM) by enhancing the CBAM utilizing the bilinear method, building upon the concept of bilinear CNN. Moreover, the Channel Attention Module (CAM) of CBAM incorporates the feature information from both the channel and spatial dimensions to boost its performance. In order to enhance the algorithm described in this research, it is important to take into account the bilinear nature of BCNN. The first max pooling and average pooling processes are kept concurrent to retain the preservation of feature information from the original maps to the maximum degree feasible. Subsequently, the output results are combined together according to the channel. The fully connected layer in the original architecture is substituted with a 1x1 convolutional layer, and batch data processing is conducted using layer normalization. ConvNeXt employs the Gaussian Error Linear Unit (GeLU) activation function, which is identical to the one used in the Swin Transformer. CBAM replaces the ReLU activation function with GeLU, which includes stochastic regularity. The proposed CBAM configuration is shown in Fig. 7.

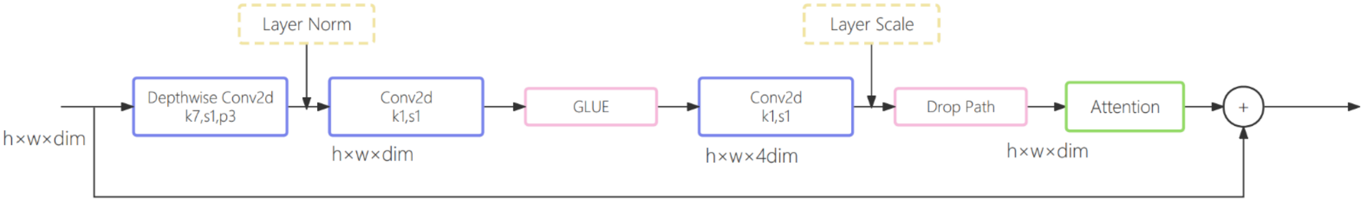


Fig. 5. Structure chart of EBA.

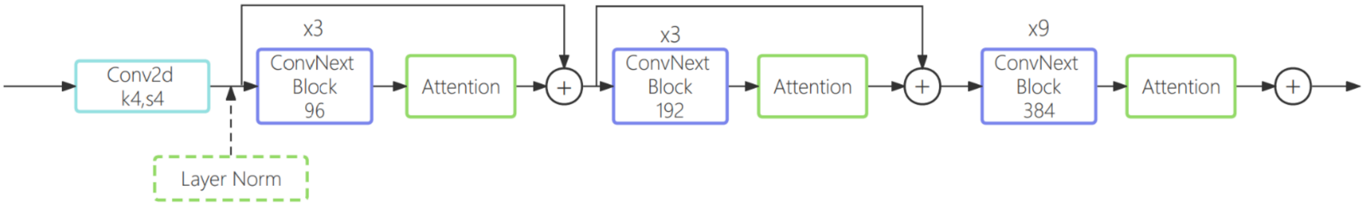


Fig. 6. Structure chart of SBA.

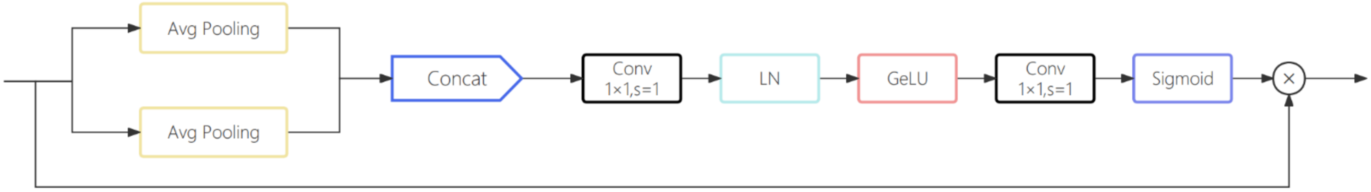


Fig. 7. Enhanced CAM structure.

Eq. (1) depicts the manner in which the network processes input data. The symbol F_{a1} denotes the map of feature generated by Enhanced CAM structure. Conv refers to the convolution operation and GeLU refers to Gaussian Error Linear Units, which are a type of activation functions. Once GeLU function applies, F becomes a tensor of shape $F \in \mathbb{R}^{(B,2C/r,W,H)}$. Upon the application of the function of sigmoid activation, F becomes a tensor of shape $F \in \mathbb{R}^{(B,C,W,H)}$. Avgpool is an abbreviation for average pooling, while Maxpool is an abbreviation for maximal pooling.

Subsequently, we incorporate bilinearity into the SAM. This work presents a new technique for extracting spatial attention, building upon existing methods. Firstly, we perform 1x1 and 3x3 convolution operations separately on the input feature maps. This allows model to obtain spatial features at different scales. Then apply layer normalization and the $GeLU$ activation function mapping to these two sets of features. Subsequently, individual channels are reduced in size using a convolution of 1 x 1. The two individual data channels are combined to generate spatial attention features at many scales. To create feature maps that combine spatial and channel weight features, features are multiplied by the feature map of channel attention weights. The data is then subjected to batch processing, and layer normalization is employed for two reasons. Firstly, we leverage the research findings of ConvNeXt, which utilizes layer normalization for batch processing in Transformer models. By applying layer normalization to the attention information in this paper, we align with the data processing method of ConvNeXt. Secondly, while batch normalization is widely used in CNNs and yields better results with larger batch sizes, Nevertheless, the hardware imposes a constraint on the batch size, which is independent of the layer normalization. When choosing the convolution kernel size, we opt for 1x1 and 3x3 convolutions. The 1x1 convolution not only allows us to obtain spatial features at different scales but also reduces the complexity and computational cost comprising the full network. Additionally, most of GPUs have implemented optimized algorithms for performing 3x3 convolutions, further improving efficiency. Based on the considerations mentioned above, we have optimized and modified the SAM structure accordingly. The improved SAM structure is illustrated in Fig. 8. The network structure's handling of input data is outlined in Eq. (2) to Eq. (4).

$$\boldsymbol{F_{a2}} = Sigmoid(\boldsymbol{F_1} \otimes \boldsymbol{F_2}) \quad (2)$$

$$\boldsymbol{F_1} = Conv_{1\times1}(GeLU(LN(Conv_{1\times1}(\boldsymbol{F})))) \quad (3)$$

$$\boldsymbol{F_2} = Conv_{1\times1}(GeLU(LN(Conv_{3\times3}(\boldsymbol{F})))) \quad (4)$$

F_{a2} refer to the map of feature generated by the structure of SAM. $Conv_{1\times1}$ refers to convolution using a 1 × 1 kernel. $Conv_{3\times3}$ refers to convolution using a 3 × 3 kernel. Once the $GeLU$ activation function applies, F has dimensions of $F \in \mathbb{R}^{(B,C/r,W,H)}$. Upon the application of the sigmoid activation function, F has dimensions of $F \in \mathbb{R}^{(B,1,W,H)}$.

In previous studies, attention mechanisms were primarily utilized to allocate weights solely between channel attentions, neglecting the consideration of spatial attention. This paper introduces a new and unique bilinear attention mechanism that consists of two separate branches. The first branch is tasked with supplying channel weights, and the secondary branch is dedicated to producing spatial weights. The initial CBAM process utilized a sequential processing approach. However, as we sought to enhance the algorithm's performance for fine-grained classification tasks, it became crucial for the network to accurately extract attention features. Consequently, we replaced the sequential processing with parallel processing to simultaneously train both channel and spatial attention. This approach ensures that multi-scale features of the target are obtained while preserving sufficient semantic information. For a detailed illustration of the specific attention network structure, please refer to Fig. 9.

The network indicated above demonstrates the processing of input data as depicted in Eq. (5):

$$F_{\text{out}} = (F_{a1} \otimes F_{\text{in}}) \otimes F_{a2} \oplus F_{\text{in}} \quad (5)$$

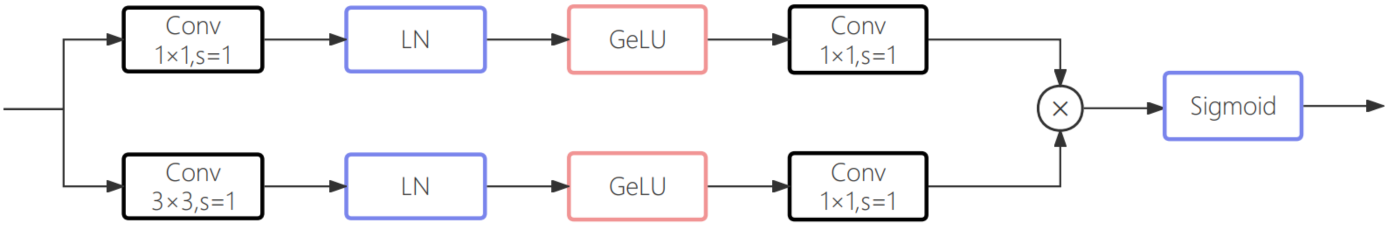


Fig. 8. Enhanced SAM structure.

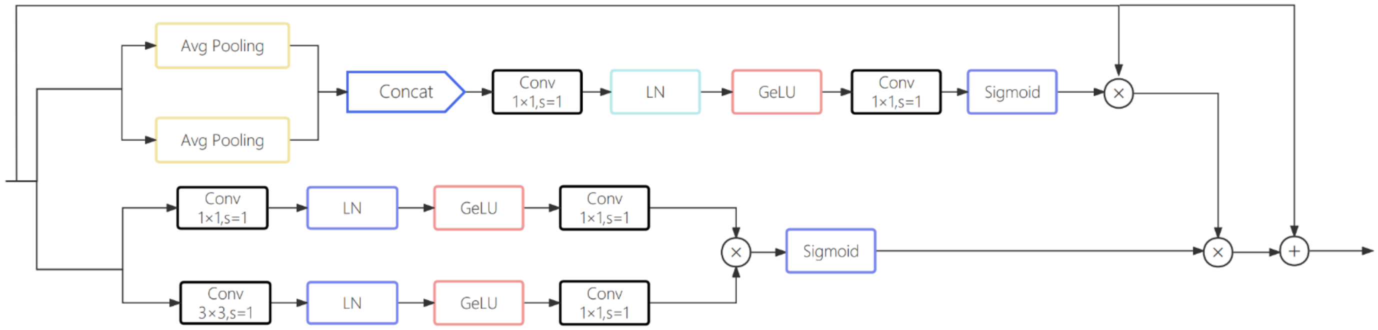


Fig. 9. Enhanced CAM structure.

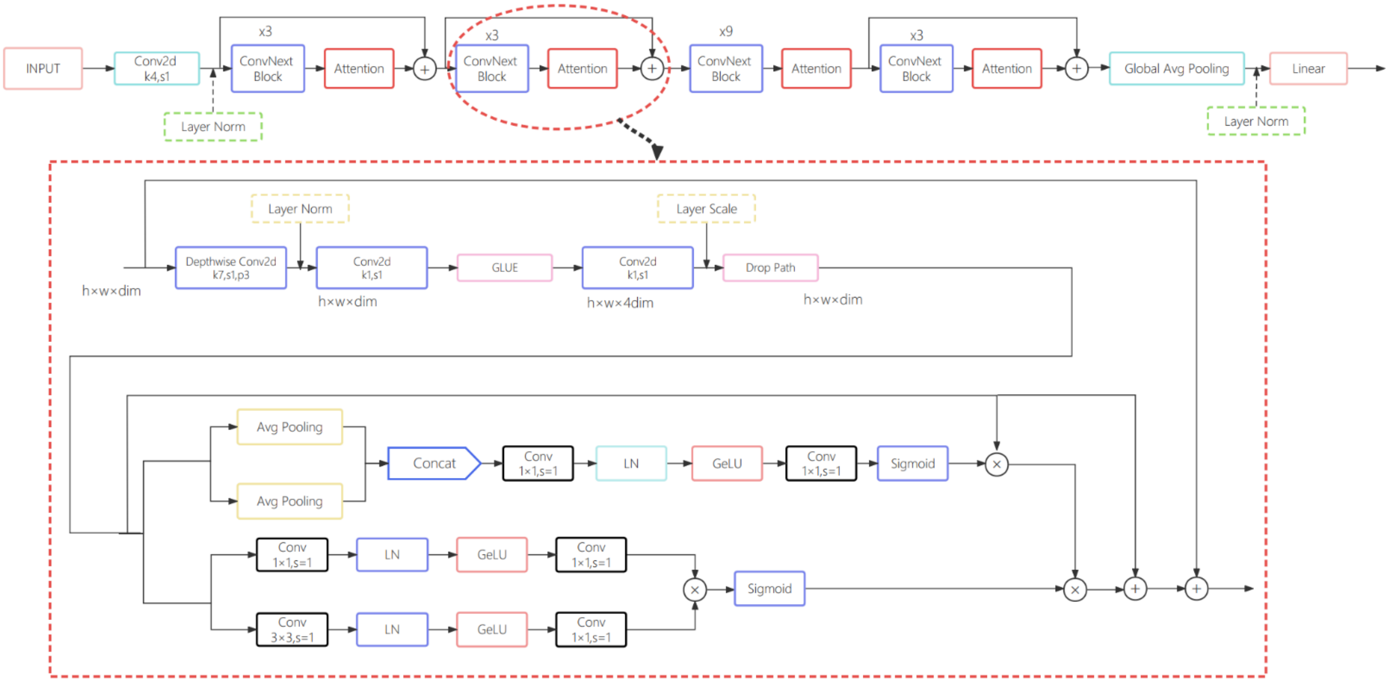


Fig. 10. ConvNeXt with BLCBAM model structure.

F_{a1} refer to the map of feature produced by CAM, F_{a2} refer to the map of feature produced by SAM, F_{in} refer to the input of BLCBAM map of feature, and F_{out} refer to the output of BLCBAM map of feature.

3) ConvNeXt Based on multiscale bilinear attention mechanism: Utilizing the proposed EBA, SBA, and multi-perspective attention framework, this paper presents a ConvNeXt model framework that incorporates bilinear attention. Through the training process, we not only extract multi-scale attention features from the feature maps iteratively but also effectively capture the fine-grained characteristics of the target.

Although ConvNeXt-Tiny already employs minimal parameters, the downsampling layers in its structure can still result in significant information loss after the conversion to images. Hence, each downsampling layer in the model is removed to preserve more information. This step enhances the information extraction ability of the embedded attention mechanism. Furthermore, considering that two-dimensional convolution can increase the dimensions of the data but may lead to information loss during convolution, the initial convolution stride in the relevant convolution modules of the original model is set to 1. This effectively strengthens the convolutional neural network's perception of the data.

Fig. 10 showcases how the proposed model structure seamlessly integrates the aforementioned components into the ConvNeXt network. As the network size and depth increase, the addition of attention modules theoretically enhances the accuracy of fine-grained classification.

IV. Implementation

In this section, the paper primarily introduces the experimental setup and implementation of the dataset, as well as presents the experimental results.

A. Dataset

To ascertain the viability of the model proposed in this paper, we conducted validation using meticulously labeled datasets obtained from the University of New Brunswick (UNB).

The datasets consist of the "ISCX VPN-nonVPN traffic dataset" (ISCX-VPN) and the "ISCX Tor-nonTor dataset" (ISCX-Tor). The primary data categorization of this dataset is displayed in Table I.

TABLE I. THE ORIGINAL DATA TYPES OF THE DATASET

Traffic Class	Application
Chat	ICQ,AIM,Skype,Facebook,Hangouts
Email	SMPT,POP3,IMAP
File transfer	Skype,FTPS,SFTP
VOIP	Facebook,Skype,Hangouts,Voipbuster
P2P	Torrent
Streaming	Viemo,Youtube,Netfilx,Spotify
VPN-Chat	ICQ,AIM,Skype,Facebook,Hangouts
VPN-Email	SMPT,POP3,IMAP
VPN-File	Transfer Skype,FTPS,SFTP
VPN-VOIP	Facebook,Skype,Hangouts,Voipbuster
VPN-P2P	Bittorrent
VPN-Streaming	Viemo,Youtube,Netfilx,Spotify

B. Data Classification Setting

After data processing, a comprehensive dataset containing all the classes was obtained for analysis in this paper. However, careful observation revealed an inherent imbalance within the dataset, which could potentially impede the overall classification performance. To address this concern, we undertook the task of reclassifying the data labels and subsequently created a balanced dataset. This was achieved by implementing the random undersampling method, which aims to maintain an equitable distribution of initial samples across each category.

The random undersampling technique functions by reducing the number of samples in the majority class, thereby rectifying the imbalance. Specifically, for the samples belonging to the majority class, random undersampling randomly selects a subset of them, ensuring that the sample size of the majority class approximates or equals that of the minority class. By adopting this approach, the model is prevented from exhibitingan excessive bias towards the majority class, consequently improving the overall classification performance. The utilization of random undersampling as a means to address the imbalanced dataset is an effective strategy that has been widely employed in various studies. Its application in this paper serves to mitigate the potential pitfalls associated with imbalanced data, ultimately enhancing the model's classification capabilities. The data classification after filtering is shown in Table II.

This paper encompasses the design of two distinct multi-class classification problems, each serving a specific purpose within the research framework:

Traffic category recognition: The primary objective of this task is to accurately identify and classify traffic based on three encryption techniques: nonVPN, VPN, and Tor. To facilitate this, the paper generated three multi-class datasets utilizing the processed data.

Application recognition: In order to assess the model's generalization capabilities, a dedicated application dataset was devised. The model underwent training on this dataset using transfer learning techniques, followed by fine-tuning using a limited amount of data from other protocols. This particular method sought to assess the model's competence in identifying various applications.

As previously mentioned, meticulous efforts were made to create filtered and balanced datasets for each sub-problem. Subsequently, the data was divided into a 90% training set and a 10% test set ratio, ensuring a robust evaluation of the model's performance.

C. Evaluation Metrics

In this paper, the evaluation of model performance is based on two key metrics: accuracy and recall. Accuracy is defined as the ratio of correctly predicted samples to the total number of predicted samples. It provides an overall measure of the model's correctness in predicting the class labels. On the other hand, recall assesses the model's ability to accurately identify positive samples. It quantifies the proportion of true positive samples correctly identified by the model. The formal definitions are as follows:

$$\text{Accuracy } = \frac{\sum_{i\in \text{ classes}} TP_i}{\sum_{i\in \text{ classes}} (TP_i+FP_i)} \tag{6}$$

$$\text{Recall } = \frac{TP}{TP+FN} \tag{7}$$

These evaluation metrics provide valuable insights into the model's overall accuracy and its ability to correctly identify positive samples. They are crucial in assessing the model's performance across different classification tasks.

TABLE II. THE FILTERING DATA TYPES OF THE DATASET

	VPN	Non-VPN	Non-Tor
Chat	AIM, Facebook, Hangouts,ICQ, Skype	AIM, Facebook, Hangouts,ICQ, Skype	-
Email	Email	Email	-
File Transfer	FTP,SFTP,Skype	FTP, SFTP, SCP	FTP,STP,POP,Skype
Streaming	Vimeo, Youtube	Hangouts,Netflix,Skype,Spotify,Vimeo,Youtube	Spotify,Vimeo, Youtube,Flash,Youtube,HTML5
VoIP	Facebook,Hangouts,Skype,Voipbuster	Facebook,Hangouts,Skype,Voipbuster	Facebook,HangoutsSkype
P2P	-	-	p2p_multipleSpeed, p2p_vuze
Browsing	-	-	Firefox, Chrome

TABLE III. THE EXPERIMENTAL OUTCOMES OF DEEP LEARNING TECHNIQUES FOR TRAFFIC IDENTIFICATION

Paper	Dataset	Method	Classification target	Accuaracy	Recall
This paper	ISCXVPN2016-NonVPN	the proposed method	Normal application traffic (10-category)	98.62%	98.95%
	ISCXVPN2016-VPN			97.43%	98.21%
	ISCXTor2016-NonTor			98.33%	98.61%
[24]	ISCXVPN2016	CAE2/ CNN	VPN and Non-VPN traffic / Traffic characterization	93.34%/92.9%	92.77%/93.5%
[25]	ISCXVPN2016	SAM(Attention Method)	Application protocol/Normal application Classification	98.62%/98.7%	98.65%/99.1%
[26]	ISCXVPN2016	PERT(Transformer)	Normal application traffic	93.27%	93.22%

D. Experimental

In this section, we have made necessary adaptations to the original model to address the specific requirements posed by smaller datasets. The original base structure, designed for classification tasks on larger datasets, needed adjustments to accommodate the significantly smaller data provided in this paper. Modifications were made to the initial convolutional parameters, and a stacking times ratio of 1:1:3:1 was implemented to prevent overfitting and safeguard the integrity of experimental accuracy.

Furthermore, to augment the model's perceptual prowess and maximize the retention of vital image information, while simultaneously safeguarding essential details, we made adjustments to the convolutional parameters and downsampling layers, as explicated in the preceding sections.

The primary objective of these adaptations is to make the model more adept at handling smaller datasets while improving its ability to perceive and interpret images. By doing so, we aim to enhance the model’s overall classification performance and ensure its suitability for the task at hand.

This paper presents a performance comparison between the improved structure and the original structure, as shown in Table III. The improved structure exhibits significant improvements in both accuracy and recall on non-VPN and VPN data. Particularly, the improved structure shows a more pronounced improvement on encrypted data compared to non-encrypted data. The average improvement rate of the improved structure is 14.75%.

For the adjusted model structure, the BLCBAM attention module is incorporated and used for classifying both encrypted and non-encrypted traffic. The traffic data is processed and transformed into images, which are then fed into the model for classification. The specific experimental results can be seen in Table IV.

TABLE IV. PERFORMANCE IMPROVEMENT AFTER ADAPTATION

	Non-VPN		VPN	
	Ac(%)	Re(%)	Ac(%)	Re(%)
Original version	78.53	77.95	82.35	81.32
Improved version.	94.35	94.67	95.14	95.3
promotion	15.8	16.7	12.6	13.9

The results in Table IV demonstrate that this paper achieves a recognition accuracy of 98.62% for non-VPN traffic, 97.43% for encrypted VPN traffic, and 98.33% for Tor traffic on this dataset. Compared to previous studies, the improved model in this paper shows a 5% improvement relative to the model proposed by Draper-Gil et al. [24] that uses CAE and CNN. Furthermore, compared to the model proposed by He H Y et al. [26] that combines Transformer and transfer learning, the model in this paper exhibits more accurate performance.

TABLE V. ADAPTATION RESULTS FOR SINGLE APPLICATION RECOGNITION

Classification	Accuracy (%)			
Chat	Training/Test	Non-VPN	VPN	Non-Tor
	Non-VPN	98.6	82.9	-
	VPN	75.7	96.4	-
	Non-Tor	-	-	-
Email	Training/Test	Non-VPN	VPN	Non-Tor
	Non-VPN	96.2	89.2	-
	VPN	69.7	98.2	-
	Non-Tor	-	-	-
File Transfer	Training/Test	Non-VPN	VPN	Non-Tor
	Non-VPN	98.8	73.9	81.6
	VPN	60.1	96.9	67.3
	Non-Tor	79.2	65.8	97.6
Streaming	Training/Test	Non-VPN	VPN	Non-Tor
	Non-VPN	99.8	71.9	70.6
	VPN	72.1	96.8	54.5
	Non-Tor	83.1	92.8	94.8
VoIP	Training/Test	Non-VPN	VPN	Non-Tor
	Non-VPN	95.6	69.4	85.2
	VPN	67.8	98.1	80.4
	Non-Tor	89.1	83.8	93.3

The comparison of traffic classification using different algorithms has been presented in the Table VI. The C4.5 machine learning algorithm exhibited suboptimal performance in accurately identifying both VPN and non-VPN data streams, with a precision value below 85%. Our proposed strategy in the domain of deep learning has exhibited a substantial enhancement in accuracy. Compared to algorithms utilizing a single model, the approach presented in this paper achieved an approximately 5% enhancement in VPN accuracy and a substantial 12% improvement in Non-VPN accuracy. Regarding the composite model approach, the method proposed in this paper exhibits similar performance in the VPN domain, while achieving an approximately 10% improvement in the Non-VPN domain. It is evident that, in comparison to prior research, the proposed method in this paper has achieved noteworthy advancements in traffic classification.

In certain real-world scenarios, it is necessary to identify whether specific protocols or applications exist within a large volume of network traffic. This paper conducts experiments specifically for this situation and trains the model using different encryption techniques to identify specific traffic categories. Transfer learning is employed using a small amount of data for the target protocols to improve the model's transferability. The results of the experiment are displayed in the Table V.

TABLE VI. OVERALL RESULT ON VARIOUS ALGORITHMS

Method	**VPN**		**Non-VPN**	
	Accuracy(%)	**Recall(%)**	**Accuracy(%)**	**Recall(%)**
C4.5 [27]	78.2	81.3	84.3	79.3
ID CNN [17]	92	95.2	85.8	85.9
SAE+1DCNN [28]	97.8	96.3	86.7	88.8
This Paper	97.4	98.2	98.6	98.9

By taking the average of the results for each encryption technique and considering only the cases where the test set has the same traffic category and encryption technique as the training set, the average accuracy obtained in this paper is as follows: 97.8% for non-VPN traffic, 97.28% for VPN traffic, and 94.9% for Tor traffic. Considering the overall average accuracy based on the above values, the overall average accuracy obtained in this paper is 96.66%. This paper has achieved significant success in describing and identifying internet traffic categories transmitted through different encryption techniques.

V. CONCLUSION

This paper presents a novel approach for traffic classification, employing ConvNeXt and a bilinear attention mechanism, with the goal of automatically extracting and analyzing traffic features to achieve accurate traffic characterization and application classification. The proposed method entails the conversion of data traffic into image representations, followed by the utilization of the ConvNeXt framework model, coupled with the bilinear attention mechanism, to enhance the model's perception of image features and optimize classification accuracy. Through extensive experimentation on various meticulously designed datasets, this paper thoroughly examines the performance of the proposed method and conclusively demonstrates its remarkable efficacy in accurately classifying diverse internet traffic categories.

REFERENCES

[1] Cotton, Michelle, Lars Eggert, and Dr. Joseph D. Touch. "Request for comments: No. 6335 Internet Assigned Numbers Authority (IANA) Procedures for the Management of the Service Name and Transport Protocol Port Number Registry." 2011.

[2] Moore, A. W., and K. Papagiannaki. "Toward the accurate identification of network applications." Passive and Active Network Measurement: 6th International Workshop, PAM 2005, Boston, MA, USA, March 31-April 1, 2005. Proceedings 6. Springer Berlin Heidelberg, 2005. 41-54.

[3] Moore, A. W., and D. Zuev. "Internet traffic classification using Bayesian analysis techniques." Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems. 2005. 50-60.

[4] Liu, Z., H. Mao, C. Y. Wu, et al. "A convnet for the 2020s." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. 11976-11986.

[5] P. Khandait, N. Hubballi, B. Mazumdar, "Efficient keyword matching for deep packet inspection based network traffic classification," in: Proceedings of the 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), IEEE, 2020, pp. 567–570.

[6] X. Wang, J. Jiang, Y. Tang, B. Liu, X. Wang, "Strid2fa: scalable regular expression matching for deep packet inspection," in: Proceedings of the 2011 IEEE International Conference on Communications, ICC, 2011, pp. 1–5.

[7] S. Fernandes, R. Antonello, T. Lacerda, A. Santos, D. Sadok, "Slimming down deep packet inspection systems," in: Proceedings of the IEEE INFOCOM Workshops, 2009, pp. 1–6.

[8] K.L. Dias, M.A. Pongelupe, W.M. Caminhas, L. de Errico, "An innovative approach for real-time network traffic classification," Computers & Networks 158 (2019), 143–157.

[9] J. Cao, D. Wang, Z. Qu, H. Sun, B. Li, C.-L. Chen, "An improved network traffic classification model based on a support vector machine," Symmetry 12 (2) (2020), 301.

[10] Y. Wang, Y. Xiang, J. Zhang, S. Yu, "A novel semi-supervised approach for network traffic clustering," in: Proceedings of the 2011 5th International Conference on Network and System Security, 2011, pp. 169–175.

[11] Höchst J, Baumgärtner L, M. Hollick, B. Freisleben, "Unsupervised traffic flow classification using a neural autoencoder," in: Proceedings of the 2017 IEEE 42nd Conference on Local Computer Networks, LCN, 2017, pp. 523–526.

[12] T. Bakhshi, B. Ghita, "On internet traffic classification: a two-phased machine learning approach," Journal of Computer Networks and Communications 2016 (2016), 21.

[13] F. Noorbehbahani, S. Mansoori, "A new semi-supervised method for network traffic classification based on x-means clustering and label propagation," in: Proceedings of the 2018 8th International C

[14] Shapira, T., and Y. Shavitt. "Flowpic: Encrypted internet traffic classification is as easy as image recognition." IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2019. 680-687.

[15] Lan, J., X. Liu, B. Li, et al. "DarknetSec: A novel self-attentive deep learning method for darknet traffic classification and application identification." Computers & Security 116 (2022): 102663.

[16] D'Angelo, G., and F. Palmieri. "Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial-temporal features extraction." Journal of Network and Computer Applications 173 (2021): 102890.

[17] Wang, W., M. Zhu, J. Wang, et al. "End-to-end encrypted traffic classification with one-dimensional convolution neural networks." 2017 IEEE international conference on intelligence and security informatics (ISI). IEEE, 2017. 43-48.

[18] Wang, W., M. Zhu, X. Zeng, et al. "Malware traffic classification using convolutional neural network for representation learning." 2017 International conference on information networking (ICOIN). IEEE, 2017. 712-717.

[19] Maonan, W., Kangfeng, Z., Ning, X., et al. "Centime: A direct comprehensive traffic features extraction for encrypted traffic classification." 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS). IEEE, 2021. 490-498.

[20] Barut, O., Luo, Y., Li, P., et al. "R1dit: Privacy-preserving malware traffic classification with attention-based neural networks." IEEE Transactions on Network and Service Management (2022).

[21] Liu, X., You, J., Wu, Y., et al. "Attention-based bidirectional GRU networks for efficient HTTPS traffic classification." Information Sciences 541 (2020): 297-315.

[22] Xiao, X., Xiao, W., Li, R., et al. "EBSNN: extended byte segment neural network for network traffic classification." IEEE Transactions on Dependable and Secure Computing 19.5 (2021): 3521-3538.

[23] Woo, S., Park, J., Lee, J. Y., et al. "CBAM: Convolutional block attention module." Proceedings of the European conference on computer vision (ECCV). 2018. 3-19.

[24] Draper-Gil, G., Lashkari, A. H., Mamun, M. S. I., et al. "Characterization of encrypted and VPN traffic using time-related." Proceedings of the 2nd international conference on information systems security and privacy (ICISSP). 2016. 407-414.

[25] Xie, G., Li, Q., Jiang, Y., et al. "SAM: Self-attention based deep learning method for online traffic classification." Proceedings of the Workshop on Network Meets AI & ML. 2020. 14-20.

[26] He, H. Y., Yang, Z. G., Chen, X. N. "PERT: Payload Encoding Representation from Transformer for Encrypted Traffic Classification." 2020 ITU Kaleidoscope: Industry-Driven Digital Transformation (ITU K). 2020. DOI:10.23919/ITUK50268.2020.9303204.

[27] Lashkari, A. H., Draper-Gil, G., Mamun, M. S. I., et al. "Characterization of Tor traffic using time-based features." ICISSp. 2017. 253-262.

[28] Lotfollahi, M., Jafari Siavoshani, M., Shirali Hossein Zade, R., et al. "Deep packet: a novel approach for encrypted traffic classification using deep learning." Soft Computing, 24, 1999–2012 (2020).

Energy-Aware Clustering in the Internet of Things using Tabu Search and Ant Colony Optimization Algorithms

Mei Li[1], Jing Ai[2]*
School of Electrical Engineering and Automation, Anhui University, Hefei 230601, Anhui, China[1]
Information Engineering College, Wuhan Design Engineering College, Wuhan 430000, Hubei, China[2]

Abstract—**The Internet of Things (IoT) significantly impacts communication systems' efficiency and the requirements for applications in our daily lives. Among the major challenges involved in data transmission over IoT networks is the development of an energy-efficient clustering mechanism. Recent methods are challenged by long transmission delays, imbalanced load distribution, and limited network lifespan. This paper suggests a new cluster-based routing method combining Tabu Search (TS) and Ant Colony Optimization (ACO) algorithms. The TS algorithm overcomes the disadvantage of ACO, in which ants move randomly throughout the colony in search of food sources. In the process of solving optimization problems, the ACO algorithm traps ants, resulting in a considerable increase in the time required for local searches. TS can be used to overcome these drawbacks. In fact, the TS algorithm eliminates the problem of getting stuck in local optima due to the randomness of the search process. Experimental results indicate that the proposed hybrid algorithm outperforms ACO, LEACH, and genetic algorithms regarding energy consumption and network lifetime.**

Keywords—Internet of things; clustering; data transmission; energy efficiency; ant colony optimization algorithm

I. Introduction

The Internet of Things (IoT) envisions the seamless integration of smart devices. The IoT enables smart objects to interact among themselves, aggregate information within a network, and combine digital and physical objects to create unique experiences that meet certain end-user requirements [1, 2]. Communication takes place between people and things and between things themselves. A broad range of real-world applications has been supported by data aggregation in terms of reducing energy consumption by Wireless Sensor Networks (WSNs). For example, when it is necessary to monitor an area for a particular purpose continuously, multiple sensor nodes may be placed in the area [3, 4]. The information collected from the sensors is gathered, summarized, and transmitted to the base station to address specific questions. When the surrounding environment remains relatively stable, individual sensor data may show a high level of temporal correlation, indicating that two successive values are unlikely to differ significantly. Due to the high energy consumption of such an application, it becomes necessary to minimize transmission sensor data with no changes [5].

The field of Artificial Intelligence (AI) has been rapidly advancing in recent years, with applications in various domains such as finance, healthcare, and the IoT. One area of interest is the use of deep learning-based models for stock price prediction, as discussed in research [6]. Another area of research is the design of efficient data collection methods for IoT networks, such as the use of unequal sized cells based on cross shapes proposed by Taami, et al. [7]. In the healthcare domain, Soleimani and Lobaton [8] proposed a phase-based interpretability and multi-task learning approach to enhance inference on physiological and kinematic periodic signals. In the energy sector, Bagheri, et al. [9] developed a data conditioning and forecasting methodology using machine learning for production data on a well pad. In the field of wireless communications, Webber, et al. [10] proposed a probabilistic neural network for predicting idle slot availability in WLANs, while in other studies, they explored the use of machine learning for human activity recognition [11], vaccine candidate prediction [12], network slicing [13], and green smart cities [14]. These studies demonstrate the potential of AI and machine learning techniques to address various challenges and opportunities in different domains.

In many cases, IoT devices operate on short-life and non-rechargeable batteries. These batteries need to be replaced periodically, which can be costly and time-consuming [15]. Furthermore, these batteries can potentially be a source of pollution if not disposed of properly. Recharging or replacing these batteries can be difficult and expensive [16, 17]. As a result, it is important to design IoT devices with energy-efficient components that can run on minimal power for extended periods. The ability to process, aggregate, and transmit data in an energy-efficient manner plays a vital role in IoT applications. By using low-power components, such as sensors and microprocessors, IoT devices can operate efficiently and cost-effectively [18]. This allows these devices to run on minimal power and, in turn, prolongs their lifespan. Wireless communications consume more energy than processing in internet-based systems. Thus, achieving a mechanism for transmitting data between sources and destinations is an important challenge in IoT. Clustering in IoT is crucial to the appropriate transmission of data. This process involves grouping devices into clusters and assigning them cluster heads to enhance resource utilization [19, 20].

In the clustering process in IoT-based, sensor nodes are initially deployed in a network. The system performance is improved by forming a cluster of nodes. The optimal CH is determined for each cluster under different performance metrics. The CHs collect data packets from non-Ch nodes and forward them to the IoT base station. A cloud server will be used to store the data obtained from the base station. During the data processing step, various analytics approaches are employed to eliminate noisy and inconsistent data. The outcome will be available to the end users upon completion of the process. Cluster head selection objectives include monitoring and managing network lifetimes, energy consumption, node failures, load balancing, and network resources. Various clustering mechanisms have been proposed in the literature, including heuristics, metaheuristics, and fuzzy-based approaches. A majority of heuristic algorithms are designed to minimize the number of clusters. In meta-heuristic clustering algorithms, the distance between devices and the remaining energy is considered key performance indicators, whereas data volume and the number of one-hop neighbors are not considered. Moreover, fuzzy-based algorithms rely on assumptions, and validation and verification need extensive tests.

In this paper, we propose a novel approach to CH selection by combining ACO and TS algorithms. Our method integrates the strengths of both algorithms synergistically. Specifically, it leverages the TS algorithm's robust search capabilities and rapid convergence to effectively address local optima issues commonly encountered with ACO. This fusion of ACO and TS enhances the efficiency and effectiveness of CH selection in IoT networks, ultimately contributing to the overarching goal of improving network performance while conserving energy and extending the network's lifespan. Through this research, we aim to provide a practical and innovative solution to the challenges associated with CH selection in IoT, offering a promising avenue for optimizing IoT network operation and sustainability.

II. Related Work

Mohseni, et al. [19] proposed a cluster-based routing strategy called CEDAR by combining the fuzzy logic system with the Capuchin search algorithm. Clustering is applied to both intra-cluster and extra-cluster routing. In this strategy, nodes in the network are clustered to reduce energy consumption, which is a significant benefit to IoT devices. Packets are routed between nodes within each cluster. Nodes can adapt to changing network conditions with the fuzzy logic system, and packets are routed efficiently with the Capuchin search algorithm. CEDAR performed better than comparative approaches in terms of energy consumption, network delay, and network lifetime based on simulation results. Based on the Sailfish optimization algorithm, Sankar, et al. [21] proposed a new method for selecting CHs and forming clusters. NS2 simulator is used for the simulation. This study compares the efficacy of SOA with hierarchical clustering-based, optimized particle swarm optimization, and improved ACO. In the simulation, it was demonstrated that the proposed SOA increases network life and reduces node-to-sink delays.

A new clustering method has been proposed by Yarinezhad and Sabaei [22] for balancing traffic loads in IoT-enabled WSNs. A 1.2 approximation algorithm is employed in the proposed clustering method. A new energy-aware routing algorithm is introduced to enable data packets to be transmitted from the CHs to their destinations. This algorithm allows data packets to be distributed among several nodes in the vicinity of the destination by segmenting the area properly. According to test results, the proposed clustering algorithm is not only suitable for large-scale IoT-enabled WSNs but also demonstrates superior performance over other algorithms of a similar nature. Senthil, et al. [23] proposed a new clustering method based on the particle swarm optimization (PSO) algorithm. Particles in the PSO represent candidate solutions and tend to move through their solution space at varying speeds (in several directions). Experimental results demonstrate that the proposed method optimizes the clustering process and achieves energy efficiency. In addition to reducing end-to-end delays and packet loss rates, the lifespan network and cluster count have been improved.

Maheswar, et al. [24] presented a cluster-based backpressure routing (CBPR) scheme to extend network lifetime and improve data transmission reliability through energy load balancing. Depending on the energy level and distance to the sink node, the CBPR scheme decides which cluster head to elect for each cluster of the sensor node. Additionally, the proposed CBPR routing scheme utilizes a highly robust data aggregation algorithm to prevent redundant data packets from circulating throughout the network. For data packet queuing and route selection, the backpressure scheduling machine is utilized, allowing it to determine the next-hop sensor node based on the queue lengths of sensor nodes. CBPR routing scheme has been evaluated extensively through extensive simulations, compared with those of other well-known routing schemes, including Information Fusion Based Role Assignment and Data Routing for In-Network Aggregation, in terms of throughput, energy consumption, and packet delivery.

Aravind and Maddikunta [25] introduced a cluster-based routing protocol for IoT based on a Self-Adaptive Dingo Optimizer with Brownian Motion (SDO-BM) algorithm to select optimum CHs under parameters including QoS, trust, overhead, delay, distance, and energy. The proposed protocol showed promising results in terms of energy consumption, latency, and packet delivery rate. It also had the ability to self-adapt to changing network conditions, making it a reliable and efficient routing protocol for IoT networks. This approach effectively uses an alternative CH, thus reducing the impact of a node failure. Additionally, it also helps conserve energy since it avoids the need to re-elect a new CH. By using this protocol, networks can become more reliable and efficient.

Energy-efficient clustering in the IoT networks faces several notable challenges. One significant challenge is the issue of long transmission delays. Many existing clustering algorithms struggle to strike a balance between data transmission efficiency and the need to conserve energy. As a result, data packets often experience delays, hindering real-time applications critical in IoT, such as remote monitoring and control. Another challenge lies in load distribution. Current approaches often suffer from imbalanced load distribution among CHs in the network. This imbalance can lead to premature energy depletion of some CHs, leaving parts of the network vulnerable and causing network degradation. Ensuring a fair and efficient distribution of responsibilities among CHs is a complex problem that needs to be addressed effectively. Furthermore, limited network lifespan remains a persistent challenge. The energy resources of IoT devices are inherently constrained, making it crucial to maximize network longevity. Many existing algorithms do not adequately optimize energy consumption, leading to shortened network lifespans. As IoT deployments continue to grow, addressing this issue becomes paramount to sustainability and cost-effectiveness.

Current approaches in energy-efficient clustering for IoT networks exhibit several limitations that hinder their effectiveness. One common limitation is the tendency to converge to local optima. Many clustering algorithms face challenges in escaping local optima due to their exploration-exploitation trade-off. This limitation can prevent the algorithms from discovering more energy-efficient solutions. Additionally, current approaches often lack adaptability to dynamic network conditions. IoT environments are dynamic, with nodes joining and leaving the network regularly. Many clustering algorithms struggle to adapt to these changes, resulting in suboptimal performance and the need for manual adjustments. Moreover, the scalability of current methods is a concern. As IoT networks grow in size and complexity, existing algorithms may struggle to handle the increased computational demands, potentially leading to performance degradation. Furthermore, the lack of a unified evaluation framework makes it challenging to compare the performance of different clustering algorithms objectively. This fragmentation hinders the identification of the most suitable algorithm for specific IoT deployment scenarios, limiting the practical applicability of current approaches. Addressing these limitations is essential to advance the field and provide more robust and adaptable clustering solutions for IoT networks.

III. Proposed Method

IoT networks consist of numerous nodes with varying capabilities in terms of power, processing, and storage. Sensor nodes continuously sense the network's data and relay it to the base station. Since the base station is overloaded for the aforementioned reasons, IoT sensor nodes may fail, redundant data may be generated, and temperature levels may increase. As a solution to these issues, existing nodes are grouped into clusters, with a CH chosen for each cluster based on its optimal performance. CHs are selected according to several factors, including the distance between the base station and the CH, the delay in passing the nodes to the base station, the network load, and the temperature and energy of the nodes. By selecting a suitable CH, the lifetime of the IoT network will be extended.

Traditionally, WSNs are characterized by terms such as distance, delay, and energy. Nevertheless, when choosing a CH for IoT, the load of the network and the temperature are also taken into account in addition to the above criteria. Consequently, a node with maximum energy, a minimum proximity to the base station, a minimum delay, a minimum load, and a low temperature is selected as a CH to optimize network performance. An optimal fitness function is illustrated in Eq. (1) to enhance the efficiency and stability of a network.

$$\mathrm{F} = \mathrm{a1} \times \mathrm{Load} + \mathrm{a2} \times \mathrm{Temp} + \mathrm{a3} \times (1 - \mathrm{Delay}) + \mathrm{a4} \times (1 - \mathrm{Distance}) + \mathrm{a5} \times \mathrm{Energy} \quad (1)$$

where, a1, a2, a3, a4, and a5 represent weighted parameters, and their sum equals one.

A. Load and Temperature

Choosing the optimal CH requires minimal load and temperature on sensor nodes. The Xively IoT platform monitors the performance of the nodes by collecting load and temperature data. As a Google IoT platform, Xively [26] connects, manages, and engages products in milliseconds across millions of connections. Scalability and performance are two of Xively's key features. Environmental monitoring, home automation systems, remote control systems, and building management systems are some application scenarios that can be considered. Data pertaining to the load and temperature of the sensor nodes are fed into Xively, and then the simulation's performance is evaluated. Load and temperature data are transmitted using the MQTT protocol.

B. Delay

An increase in network efficiency can be achieved by transmitting data packets in a limited period. In order to measure the delay in packet transmission to the destination, two factors are taken into account: transmission delay (Tt) and propagation delay (Tp). Tt is the time taken to send the data packet from the source to the destination. Tp is the time taken for the packet to travel from the source to the destination. Both delays must be considered when measuring network efficiency. In Eq. (2), the objective function for measuring the latency time for packets to be transferred from the CH to the base station is shown.

$$\mathrm{Delay} = \frac{\mathrm{Max} \sum_{\mathrm{t}=1}^{\mathrm{TCL}} \mathrm{CL}_{\mathrm{t}}}{\mathrm{A}} \quad (2)$$

C. Distance

The objective function for the distance between the sensor node and CH and BS is expressed in Eq. (3). In the case of trivial distances from the node to the BS, the optimal CH is chosen.

$$\mathrm{Distance} = \sum_{\mathrm{x}=1}^{\mathrm{X}} \sum_{\mathrm{t}=1}^{\mathrm{TCL}} \frac{\|\mathrm{Dis}_{\mathrm{S}}^{\mathrm{x}} - \mathrm{Dis}_{\mathrm{CL}}^{\mathrm{t}}\| + \|\mathrm{Dis}_{\mathrm{CL}}^{\mathrm{t}} - \mathrm{Dis}_{\mathrm{BS}}\|}{\mathrm{M} \times \mathrm{M}} \quad (3)$$

$\|\mathrm{Dis}_{\mathrm{S}}^{\mathrm{x}} - \mathrm{Dis}_{\mathrm{CL}}^{\mathrm{t}}\|$ indicates the distance between the xth sensor node and the corresponding tth cluster head. $\|\mathrm{Dis}_{\mathrm{CL}}^{\mathrm{t}} - \mathrm{Dis}_{\mathrm{BS}}\|$ denotes the distance between the tth cluster head and the BS, while M refers to the area of sensing in meters.

D. Energy Consumption

A network's lifetime and performance are greatly influenced by residual energy. An optimal CH should be selected when node energy is at a high level. Upon transmission and reception of the data packets, CH and normal nodes' energies are revised. In Eq. (4), the actual energy of a sensor node can be determined once the packets have been passed to the cluster head. Eq. (5) shows the remaining energy available in CH. Data can be transferred to CH until the node's energy reaches zero. The energy fitness function is represented in Eq. (6). In Eq. (4), $E_{X+1}(C_S^x)$ represents energy dissipation in the regular node upon transmission to the cluster head, $E_X(C_S^x)$ represents energy dissipation in the xth node. In Eq. (5), $E_{X+1}(C_{CL}^t)$ represents the energy available in CH following the transfer of data packets from the normal node, $E(C_{CL}^t)$ indicates the energy dissipated by tth cluster head.

$$E_{X+1}(C_S^x) = E_X(C_S^x) - E(C_S^x) \tag{4}$$

$$E_{X+1}(C_{CL}^t) = E_X(C_{C:}^t) - E(C_{CL}^t) \tag{5}$$

$$\text{Energy} = \frac{1}{A}\{\textstyle\sum_{x=1}^{A} E(C_S^x)\} + \frac{1}{TCL}\{\textstyle\sum_{t=1}^{CL} E(E_{CL}^t)\} \tag{6}$$

The ACO algorithm originated from the behavior of real ants in their search for the shortest route to food. A number of cycles (iterations) are involved in the construction of the solution. A number of ants construct complete solutions in each iteration based on heuristic information and previous groups' experiences. An important aspect of the ACO algorithm is the transition of ants and pheromone updates. In order to find food, ants establish the shortest paths to reach the food source. The ACO algorithm constructs solutions given the problem data and is capable of solving discrete optimization problems. As a general rule, ants search for food sources in a random manner. When an ant finds a food source, it returns some food to its colony. During their travels along the path, they leave behind chemical substances known as pheromones.

Consequently, shorter paths are likely to contain a greater concentration of pheromone trails. Pheromone trails function as a communication mechanism between ants. The intensity of pheromone trails present on the ground is determined by the quality of the solution (food source) found on the ground. Shorter paths accumulate pheromone trails with multiple ants, leading to a higher density than longer paths. This increases the appeal of shorter paths. An evaporation rate reduces all pheromone trails over time. Meanwhile, evaporation presents an opportunity for exploration and minimizes local stalling [27, 28].

$$P_{ij}^k = \begin{cases} \dfrac{(\tau_{ij})^a(\eta_{ij})^\beta}{\sum_{m\in N_i^k}(\tau_{im})^a(\eta_{im})^\beta} & j \in N_i^k \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where, P_{ij}^k represents the probability of an ant k moving from node i to node j. Pheromone levels and heuristic information are important factors in this decision. α and β refer to the relative importance of heuristic information and pheromone concentration. τ_{ij} represents the pheromone concentration on edges i and j, η_{ij} refers to the heuristic function, and N_i^k denotes an unexplored neighborhood set. Pheromone updates can be expressed in the following manner:

$$\tau_{ij} \leftarrow \tau_{ij} + \Delta\tau_{ij}^k \tag{8}$$

Evaporation updates are provided by:

$$\tau_{ij} \leftarrow (1 - \rho)\tau_{ij} \tag{9}$$

$\Delta\tau_{ij}^k$ refers to the cost of the solution provided by ant k and ρ denotes a constant factor reduction of all pheromones.

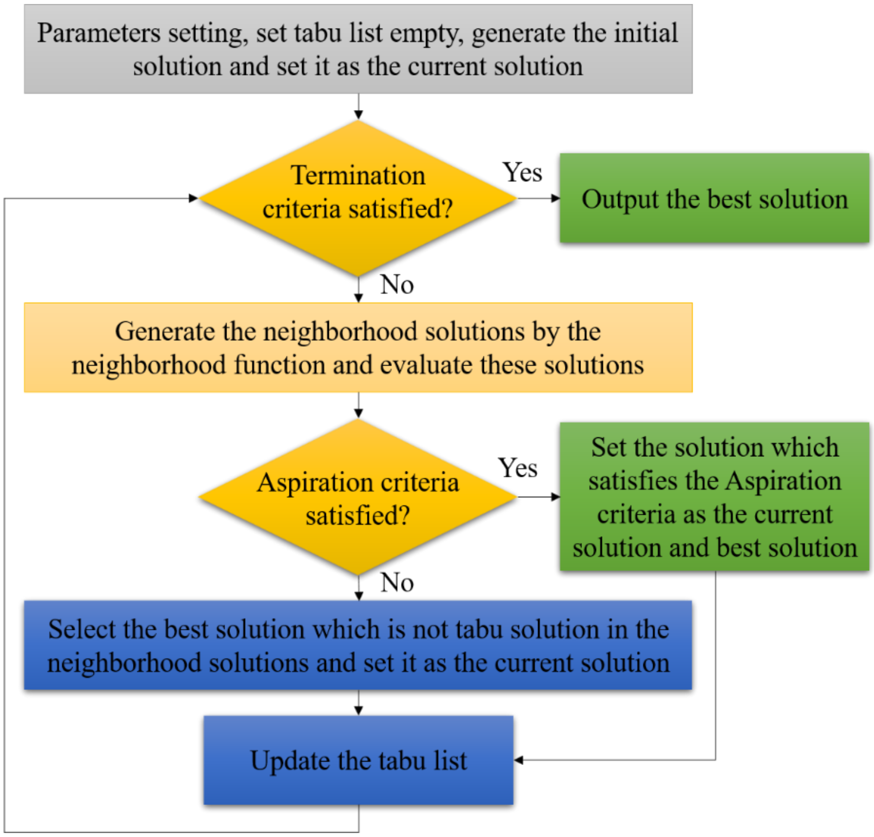

Parameters setting, set tabu list empty, generate the initial solution and set it as the current solution
Termination criteria satisfied?
Yes
Output the best solution
No
Generate the neighborhood solutions by the neighborhood function and evaluate these solutions
Aspiration criteria satisfied?
Yes
Set the solution which satisfies the Aspiration criteria as the current solution and best solution
No
Select the best solution which is not tabu solution in the neighborhood solutions and set it as the current solution
Update the tabu list

Fig. 1. Flowchart of TS.

As a meta-heuristic method, Tabu Search (TS) can be used to solve a wide range of combinatorial optimization problems. It utilizes a sequence of operators to explore the search space and generate better solutions, and it is known for its simplicity, low computational cost, and good results. Additionally, TS is a powerful tool for tackling complex problems, as it can easily be adapted to different scenarios. The final solution generally results from tracking the actions taken to transition from one solution to another. It contains a number of components, including a tabu list, neighborhood structure, move attributes, aspiration criteria, and termination conditions. TS have become recognized as a highly effective local search strategy. Fig. 1 illustrates the basic workflow of TS [29].

The disadvantage of the ACO algorithm is overcome by the TS algorithm, in which the ants move randomly in search of food sources throughout the colony. Chemical substances known as pheromones are released along the path. In the process of solving optimization problems, the ACO algorithm traps ants, which in turn results in a considerable increase in the time required for local searches. TS can be used to overcome these drawbacks. In fact, the TS algorithm eliminates the problem of getting stuck in local optima due to the randomness of the search process. This allows the ants to explore the environment better and find the best solution. The ACO algorithm uses TS to perform local searches. One of the main advantages of using the TS is that distinct parameters are used apart from the population size. ACO constitutes the core of the proposed method; however, in order to find the best solution, it employs the TS strategy when developing new solutions for every starting problem. By using distinct parameters, the TS strategy is able to generate multiple solutions to the same problem. This allows the ACO algorithm to compare and evaluate each solution, giving it the ability to determine which solution is the best one. This process is repeated until the ACO algorithm finds the optimum solution. Once the best solution is determined, the ACO algorithm terminates, and the solution is presented.

IV. Experimental Results

In this section, we present the experimental results of our proposed algorithm and compare its performance with previous algorithms, namely GA, LEACH, and ACO. The experiments were conducted using a MATLAB simulator, and the simulation data are presented in Table I. Energy consumption and network lifetime diagrams were used to illustrate the testing results and comparisons. Fig. 2 compares the algorithms in terms of energy consumption and network lifetime. Our algorithm outperforms LEACH, GA, and ACO in terms of dissipated energy, with reductions of 70%, 34%, and 17.5%, respectively, for 100 nodes. For 500 nodes, our algorithm reduces energy dissipation by 70%, 37%, and 15%, respectively, compared to LEACH, GA, and ACO. With 1000 nodes, the dissipated energy is reduced by 67%, 38.7%, and 18.3%, respectively, compared to LEACH, GA, and ACO.

Fig. 3 illustrates the comparison of the number of rounds until the last node dies versus the network size for the proposed algorithm, ACO, GA, and LEACH algorithms. Our algorithm outperforms LEACH, GA, and ACO by 15.4%, 2.3%, and 2.1%, respectively, for 100 nodes. For 500 nodes, our algorithm outperforms LEACH, GA, and ACO by 4.7%, 2.7%, and 2.6%, respectively. For 1000 nodes, our algorithm is superior to LEACH, GA, and ACO by 9.5%, 3.6%, and 1.3%, respectively. Fig. 4 compares the number of rounds until the first node drains its energy versus different network sizes. Our algorithm outperforms LEACH, GA, and ACO algorithms by 157%, 33%, and 25.3%, respectively, for 100 nodes. For 500 nodes, our algorithm is superior to LEACH, GA, and ACO algorithms by 155%, 33.8%, and 7.5%, respectively. With 1000 nodes, the proposed algorithm exceeds the performance of LEACH, GA, and ACO algorithms by 155%, 3.7%, and 6.6%, respectively.

Our proposed algorithm has been compared with several other relevant research studies in the field, and the results show that it outperforms most of them in terms of energy consumption, network lifetime, and the number of rounds until the last node dies or the first node drains its energy. One of the significant strengths of our approach is that it uses a distributed algorithm that does not require a central controller, which reduces the communication overhead and energy consumption. Additionally, our algorithm uses a dynamic threshold that adapts to the network conditions, which improves the accuracy of the algorithm and reduces the false alarms. In comparison to existing methods, our algorithm has several advantages. For example, it outperforms the LEACH algorithm in terms of network lifetime and energy consumption. The LEACH algorithm uses a fixed threshold that does not adapt to the network conditions, which results in a high false alarm rate and reduces the network lifetime. Our algorithm, on the other hand, uses a dynamic threshold that adapts to the network conditions, which reduces the false alarm rate and prolongs the network lifetime. However, our approach also has some limitations. One of the weaknesses of our algorithm is that it requires a higher computational overhead than some of the other algorithms. This is because our algorithm uses a more complex decision-making process that involves multiple parameters. Additionally, our algorithm may not be suitable for all types of wireless sensor networks, as it is designed specifically for networks with a large number of nodes and a high data rate.

TABLE I. Simulation Variables

Variable	Value
Number of nodes	100-1000
Control packet length	100 b
Data packet length	5000 b
Circuit loss	50 nJ/b
Amplification coefficient of the free space model	10 pJ/b
The initial energy of each node	0.50 J
The maximum transmission power of each node	0.005 w

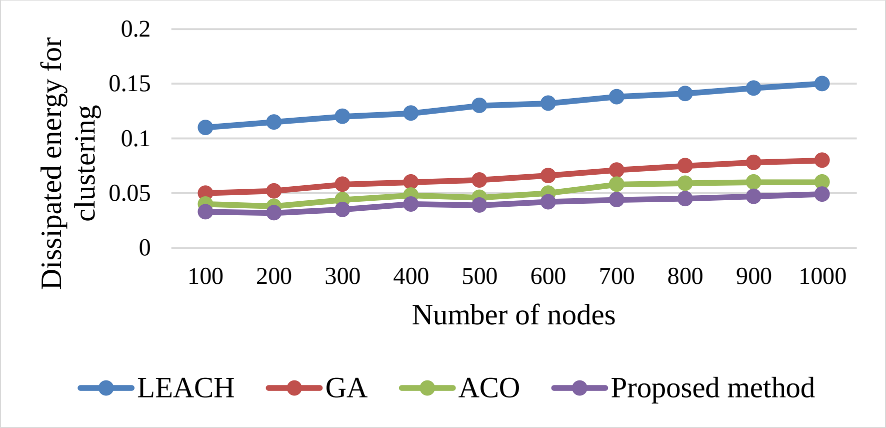


Fig. 2. Energy consumption comparison.

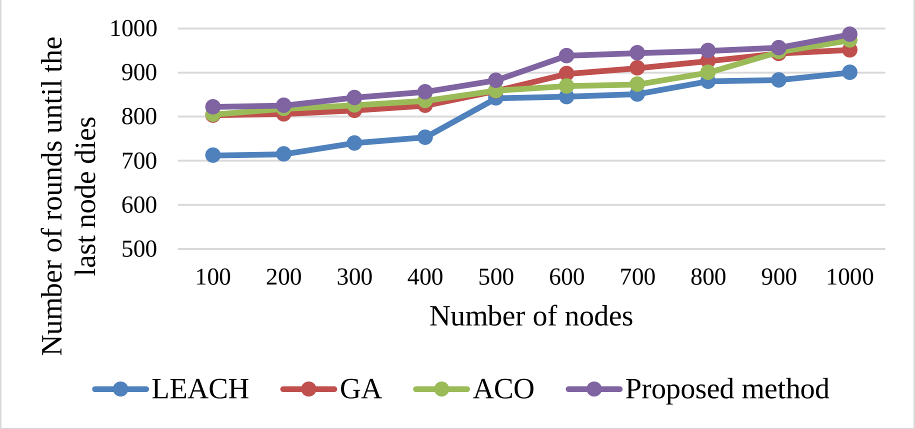


Fig. 3. Number of rounds until the last node dies.

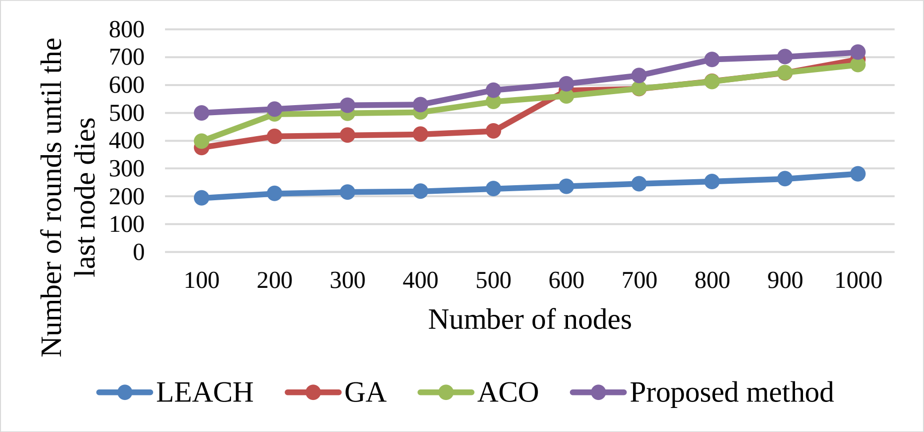


Fig. 4. Number of rounds until the first node dies.

V. Conclusion

Large-scale IoT networks collect data through sensor nodes, and the aggregated information is then sent to the next level of IoT to be processed. Considering the relatively low energy capacity of the sensor devices, IoT networks are often characterized by short battery life, resulting in a short lifespan of the network. Thus, it becomes imperative to prolong the lifespan of sensors. With clustering, collisions, interference, network redundancy, and energy consumption are reduced, and data aggregation, scalability, and network lifetime are improved. A new cluster-based routing method combining ACO and TS algorithms is presented in this paper. Using the TS algorithm, the adverse characteristics of ACO are overcome, such as the random movement of ants to find food sources in the colony. The ACO algorithm traps ants as it solves optimization problems, leading to a significant increase in the time required for local searches. TS is employed to overcome these drawbacks. Due to the randomness of the search process, the TS algorithm avoids getting stuck in local optima. Experiments have shown that the proposed hybrid algorithm performs better than ACO, LEACH, and genetic algorithms regarding energy consumption and network lifetime. There are several areas that can be explored to further improve the performance of our proposed algorithm. One possible direction is to investigate the impact of different network topologies on the performance of the algorithm. Another direction is to explore the use of machine learning techniques to optimize the parameters of the algorithm and improve its accuracy. Additionally, it may be beneficial to investigate the use of multiple thresholds to further reduce the false alarm rate and improve the network lifetime. Finally, it may be interesting to explore the use of our algorithm in other applications, such as environmental monitoring or industrial automation, to evaluate its performance in different scenarios.

REFERENCES

[1] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," Journal of Network and Computer Applications, vol. 97, pp. 23-34, 2017.

[2] A. Mehbodniya, J. L. Webber, R. Neware, F. Arslan, R. V. Pamba, and M. Shabaz, "Modified Lamport Merkle Digital Signature blockchain framework for authentication of internet of things healthcare data," Expert Systems, vol. 39, no. 10, p. e12978, 2022.

[3] M. M. Akhtar, D. Ahamad, A. S. A. Shatat, and A. E. M. Abdalrahman, "Enhanced heuristic algorithm-based energy-aware resource optimization for cooperative IoT," International Journal of Computers and Applications, pp. 1-12, 2022.

[4] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy‐efficient data fusion methods in the Internet of Things," Concurrency and Computation: Practice and Experience, p. e6959, 2022.

[5] D. Guo, "Internet of Things Based Network Security for Supply Chain Management in the Business Environment," Wireless Personal Communications, pp. 1-22, 2021.

[6] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," Frontiers in Business, Economics and Management, vol. 8, no. 2, pp. 51-54, 2023.

[7] T. Taami, S. Azizi, and R. Yarinezhad, "Unequal sized cells based on cross shapes for data collection in green Internet of Things (IoT) networks," Wireless Networks, pp. 1-18, 2023.

[8] R. Soleimani and E. Lobaton, "Enhancing Inference on Physiological and Kinematic Periodic Signals via Phase-Based Interpretability and Multi-Task Learning," Information, vol. 13, no. 7, p. 326, 2022.

[9] M. Bagheri et al., "Data conditioning and forecasting methodology using machine learning on production data for a well pad," in Offshore Technology Conference, 2020: OTC, p. D031S037R002.

[10] J. Webber, A. Mehbodniya, Y. Hou, K. Yano, and T. Kumagai, "Study on idle slot availability prediction for WLAN using a probabilistic neural network," in 2017 23rd Asia-Pacific Conference on Communications (APCC), 2017: IEEE, pp. 1-6.

[11] J. Webber, A. Mehbodniya, A. Arafa, and A. Alwakeel, "Improved Human Activity Recognition Using Majority Combining of Reduced-Complexity Sensor Branch Classifiers," Electronics, vol. 11, no. 3, p. 392, 2022.

[12] S. N. H. Bukhari, J. Webber, and A. Mehbodniya, "Decision tree based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates," Scientific Reports, vol. 12, no. 1, p. 7810, 2022.

[13] R. Singh et al., "Analysis of Network Slicing for Management of 5G Networks Using Machine Learning Techniques," Wireless Communications and Mobile Computing, vol. 2022, 2022.

[14] P. He, N. Almasifar, A. Mehbodniya, D. Javaheri, and J. L. Webber, "Towards green smart cities using Internet of Things and optimization algorithms: A systematic and bibliometric review," Sustainable Computing: Informatics and Systems, vol. 36, p. 100822, 2022.

[15] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of Medical Things Privacy and Security: Challenges, Solutions, and Future Trends from a New Perspective," Sustainability, vol. 15, no. 4, p. 3317, 2023.

[16] B. Pourghebleh, K. Wakil, and N. J. Navimipour, "A comprehensive study on the trust management techniques in the Internet of Things," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9326-9337, 2019.

[17] S. Sankar, P. Srinivasan, A. K. Luhach, R. Somula, and N. Chilamkurti, "Energy-aware grid-based data aggregation scheme in routing protocol for agricultural internet of things," Sustainable Computing: Informatics and Systems, vol. 28, p. 100422, 2020.

[18] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," Cluster Computing, pp. 1-21, 2019.

[19] M. Mohseni, F. Amirghafouri, and B. Pourghebleh, "CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic," Peer-to-Peer Networking and Applications, pp. 1-21, 2022.

[20] S. Mahmoudinazlou and C. Kwon, "A Hybrid Genetic Algorithm for the min-max Multiple Traveling Salesman Problem," arXiv preprint arXiv:2307.07120, 2023.

[21] S. Sankar, S. Ramasubbareddy, F. Chen, and A. H. Gandomi, "Energy-efficient cluster-based routing protocol in internet of things using swarm intelligence," in 2020 IEEE symposium series on computational intelligence (SSCI), 2020: IEEE, pp. 219-224.

[22] R. Yarinezhad and M. Sabaei, "An optimal cluster-based routing algorithm for lifetime maximization of Internet of Things," Journal of Parallel and Distributed Computing, vol. 156, pp. 7-24, 2021.

[23] G. Senthil, A. Raaza, and N. Kumar, "Internet of things multi hop energy efficient cluster-based routing using particle swarm optimization," Wireless Networks, vol. 27, pp. 5207-5215, 2021.

[24] R. Maheswar et al., "CBPR: A cluster-based backpressure routing for the internet of things," Wireless Personal Communications, vol. 118, pp. 3167-3185, 2021.

[25] K. Aravind and P. K. R. Maddikunta, "Dingo Optimization Based Cluster Based Routing in Internet of Things," Sensors, vol. 22, no. 20, p. 8064, 2022.

[26] N. Sinha, K. E. Pujitha, and J. S. R. Alex, "Xively based sensing and monitoring system for IoT," in 2015 International Conference on Computer Communication and Informatics (ICCCI), 2015: IEEE, pp. 1-6.

[27] R. Sekaran, A. Kumar Munnangi, S. Rajeyyagari, M. Ramachandran, and F. Al‐Turjman, "Ant colony resource optimization for Industrial IoT and CPS," International Journal of Intelligent Systems, 2021.

[28] E. B. Tirkolaee, M. Alinaghian, A. A. R. Hosseinabadi, M. B. Sasi, and A. K. Sangaiah, "An improved ant colony optimization for the multi-trip Capacitated Arc Routing Problem," Computers & Electrical Engineering, vol. 77, pp. 457-470, 2019.

[29] V. K. Prajapati, M. Jain, and L. Chouhan, "Tabu search algorithm (TSA): A comprehensive survey," in 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), 2020: IEEE, pp. 1-8.

Machine Learning-based Secure 5G Network Slicing: A Systematic Literature Review

Meshari Huwaytim Alanazi
Department of Computer Science, Northern Border University, Arar, Saudi Arabia

***Abstract*—As the fifth-generation (5G) wireless networks continue to advance, the concept of network slicing has gained significant attention for enabling the provisioning of diverse services tailored to specific application requirements. However, the security concerns associated with network slicing pose significant challenges that demand comprehensive exploration and analysis. In this paper, we present a systematic literature review that critically examines the existing body of research on machine learning techniques for securing 5G network slicing. Through an extensive analysis of a wide range of scholarly articles selected from specific search databases, we identify and classify the key machine learning approaches proposed for enhancing the security of network slicing in the 5G environment. We investigate these techniques based on their effectiveness in addressing various security threats and vulnerabilities while considering factors such as accuracy, scalability, and efficiency. Our review reveals that machine learning techniques, including deep learning algorithms, have been proposed for anomaly detection, intrusion detection, and authentication in 5G network slicing. However, we observe that these techniques face challenges related to accuracy under dynamic and heterogeneous network conditions, scalability when dealing with a large number of network slices, and efficiency in terms of computational complexity and resource utilization. To overcome these challenges, our experimentation shows that the integration of reinforcement learning techniques with CNNs, multi-agent reinforcement learning, and distributed SVM frameworks emerged as potential solutions with improved accuracy and scalability in network slicing. Furthermore, we identify promising research directions, including the exploration of hybrid machine learning models, the adoption of explainable AI techniques, and the investigation of privacy-preserving mechanisms.**

***Keywords*—5G; accuracy; deep learning; efficiency; security; machine learning; network slicing; scalability**

I. Introduction

5G networks represent the fifth generation of mobile communication networks and offer advanced features such as high-speed connectivity, ultra-low latency, and extensive machine-type communication capabilities. These networks have become crucial infrastructure for various industries due to the increasing demand for high-speed data transmission and real-time applications like autonomous vehicles, smart cities, and Industry 4.0. However, ensuring the security of 5G networks is a significant concern due to the utilization of new technologies and protocols, the complexity of the network architecture, and the potential for new attack vectors [1]. Network slicing is a crucial element of 5G networks, enabling the establishment of multiple virtual networks on a shared physical infrastructure. This technology empowers the customization of each network slice to cater to the unique demands of diverse applications and services. However, this also introduces new security challenges, including unauthorized access, data breaches, and denial of service attacks [2]. Conventional security measures such as firewalls, intrusion detection systems (IDS), and access control mechanisms may prove inadequate in effectively mitigating these threats.

Given the critical importance of 5G networks and the growing security risks, it is imperative to develop effective security mechanisms for network slicing. Machine learning-based intrusion detection systems (IDS) have emerged as a promising approach to enhance the security of 5G networks. These systems analyze network traffic, behavior patterns, and anomalies in real-time to detect and respond to security threats promptly [3]. Nevertheless, despite the potential that machine learning-based solutions hold for enhancing the security of 5G network slicing, there remains a dearth of comprehensive and methodical research in this particular domain. Therefore, the objective of this study is to conduct a thorough literature review of the current state-of-the-art in this field. The findings of this study will provide valuable insights to researchers and practitioners, helping them understand the current research trends, identify research gaps, and develop effective security solutions for 5G networks. Ultimately, this research endeavor aims to contribute to the development of secure and reliable 5G networks, which are crucial for the success of various industries and the digital economy.

This SLR is a significant contribution in the area of machine learning-based security mechanisms for 5G network slicing. The focus of this review is to analyze the security challenges through advanced machine learning and deep learning techniques, evaluating their effectiveness in the context of 5G networks slicing security, and identifying potential solutions for scalability and efficiency issues. This study analyzes various machine learning techniques for securing 5G network slicing in the subsequent sections beginning with the background in Section II. Section III details the concept of network slicing in 5G networks, explaining its implementation, business model, and significance. Section IV focuses on the security aspects of network slicing in 5G networks, addressing various security issues and challenges. Section V describes the methodology used for the systematic literature review, including data sources, search strategies, and article selection processes. Section VI presents the analysis of the findings from the literature review, discussing the advanced machine learning techniques employed for security in 5G network slicing, and highlighting the challenges and potential solutions. Section VII concludes and summarizes the key findings of the study and suggests directions for future research.

II. Background

The swift progression of technological innovations such as the Internet of Things (IoT), augmented reality (AR), and communication systems such as vehicle-to-vehicle (V2V) and vehicle-to-everything (V2X) have created a pressing demand for substantial enhancements in network and communication infrastructure [1][4]. As a response, 5G networks have gained prominence in meeting the growing consumer demands. The introduction of 5G technology has not only opened up opportunities for innovation but has also provided enhanced reliability for both service providers and consumers, resulting in a shift towards virtualization and widespread adoption of 5G. The benefits of 5G technology include exceptional data rates that are 10 to 100 times faster, widespread coverage, heightened reliability, minimal latency, enhanced quality of service (QoS), and cost-efficient service offerings. With the continuous expansion of these services and opportunities, service providers and network operators are engaged in fierce competition to deploy 5G networks and implement network slicing within the physical network.

The 5G technology consists of three distinct services, namely enhanced mobile broadband (eMBB), massive machine type communication (mMTC), and ultra-reliable low-latency communication (URLLC).eMBB offers peak data rates ranging from 10 to 100 Gbps and achieves high mobility support up to 500 km/h while ensuring reduced power consumption through the utilization of both macro and small cells. mMTC offers long-range connectivity with minimal data rates spanning from 1 to 100 Kbps, facilitating cost-effective machine-to-machine (M2M) communication. Conversely, URLLC provides highly responsive connections across multiple devices, achieving less than 1 ms latency and an end-to-end latency of 5 ms between mobile devices and base stations. URLLC also ensures moderate data rates ranging from approximately 50 Kbps to 10 Mbps, accompanied by an exceptionally high service availability of 99.9999%, establishing it as an exceptionally dependable service. [2]. The deployment of 5G networks is poised to act as a catalyst for market expansion. As of 2020, there were already 92 commercial networks operating across 38 countries, with China accounting for 150 million 5G subscribers and South Korea having eight million. Projections from Ericsson suggest that the United States alone will witness a subscriber base of 320 million by 2025 [11]. The emergence of 5G technology compels communication service providers (CSPs) to go beyond their conventional subscriber-centric business models and position themselves as digital service providers (DSPs). This transformation enables them to fuel innovation, enhance safety, and drive productivity on a global level. Recognizing the transformative potential of 5G, the World Economic Forum identifies it as a driving force behind the fourth industrial revolution. Numerous multinational companies including Huawei, Samsung, Qualcomm, LG, Ericsson, ZTE Corp, Nokia, AT&T, NEC Corp, Cisco Systems, Verizon, and Orange are actively engaged in research and development efforts related to 5G [4] [5]. Notably, AT&T, headquartered in Dallas, covers approximately 16% of the United States, while Verizon, based in New Jersey, has successfully expanded its ultra wideband network to 31 states [6]. Qualcomm predicts that by 2035, 5G will generate a staggering USD 13 trillion in value for goods and service industries worldwide [7]. Given the connectivity capabilities of 5G, which facilitate faster data rates and connect millions of devices, transitioning to 5G is imperative to meet market demands and expectations. Numerous ongoing initiatives are harnessing the potential of 5G technology to transform various industries including robotics, healthcare, automotive, agriculture, mining, media, and fashion. These projects encompass applications such as untethered industrial robots, robotic systems for agricultural purposes, AI-assisted medical diagnosis, virtual reality (VR) for palliative care, telesurgery enabling virtual patient operations, and augmented reality (AR) smart glasses for enhanced safety [8].

Delivering the aforementioned services on conventional 4G or other legacy networks presents significant challenges. To overcome these challenges and provide network services efficiently with limited resources and minimal costs for network service providers, network slicing has emerged as a promising solution. Network slicing is a key feature of 5G technology that involves partitioning the physical network into multiple logical networks, each capable of delivering customized services based on specific applications and their requirements [9]. By leveraging the progress of virtualization in cloud computing, the physical network resources are partitioned into numerous logical or virtual networks known as "slices" in the 5G context. Each network slice functions as an autonomous virtual network with dedicated resources, traffic flows, security measures, topology, and clearly defined quality of service (QoS) parameters. These slices are isolated from each other and serve the distinct service requirements of individual subscribers. [10]. Network slicing offers flexibility and scalability by allowing various services to coexist on a shared physical network. It could adapt to evolving subscriber needs, facilitate seamless end-to-end communication, support a multi-service environment, provide on-demand network services, and incorporate multi-tenancy capabilities within the 5G ecosystem. [11][12].

III. Network Slicing

The advancement and swift development of wireless communication systems have created a need for a wide array of services, applications, and scenarios tailored to meet the specific demands of enhanced mobile broadband (eMBB), ultra-reliable and low-latency communication (uRLLC), and massive machine type communication (mMTC).For instance, eMBB applications, such as virtual reality and video streaming, demand high throughput, while uRLLC services, including autonomous driving, require low latency and minimal errors. mMTC services, catering to sensing and monitoring applications, call for high connectivity. However, the existing network architecture is inadequate to meet the diverse needs of these services.

In response to this obstacle, 5G networks employ network slicing, an approach that enables the provisioning of customized services with distinct requirements over a unified network infrastructure. Network slicing includes the partitioning of the network into distinct slices, including access, transport, and core network slices. The network slicing framework is depicted in Fig. 1, illustrating this concept [14]. The core network slice consists of both the control plane and user plane, supporting shared or dedicated functions like session management, mobility management, user plane, and policy control for various slices. Notably, industry players like Ericsson and Nokia have developed their own network slicing systems tailored to their respective sectors [19]. In the current wireless communication environment, a wide range of services with different requirements in terms of security, reliability, data rate, latency, resources, and cost have emerged. Network slicing has emerged as a solution to enable resource sharing among services, customers, and providers. It requires the establishment of multiple logical networks on a shared physical infrastructure, enabling the provision of services with distinct characteristics that can concurrently support multiple technologies. Each service is allocated dedicated resources aligned with its specific requirements, thereby enhancing overall network performance. The management of network slicing involves coordinating virtual and physical resource management components, including Network Function Virtualization (NFV), Software-Defined Networking (SDN) controllers, and orchestrators [15]. NFVs (Network Function Virtualization), which refer to cloud-based functions, specify the requirements and attributes of network slices. Meanwhile, SDN (Software-Defined Networking) controllers establish instances of network slices by connecting virtual functions through SDN networks. [16][17]. Orchestrators automate the management and configuration of resources across different domains and slices, simplifying the process of creating, deploying, and monitoring services in an automated manner. Fundamentally, network slicing entails the establishment of logical networks on top of shared physical infrastructures, partitioning them into multiple virtual networks with independent control and management capabilities. Two types of orchestrators, Service Orchestrators (SOs) and Resource Orchestrators (ROs), play a crucial role in creating and managing multiple services. Several open-source network slicing orchestrators have been developed, such as openMANO, OSM, openNFV, openFV, openBaton, ZooM, ONAP, SliMANO, OpenBaton, cloudNFV, JOX, FlexRAN and Cloudify, to efficiently manage resources in access, core, and transport networks [15][18]. Notably, Huawei has developed an end-to-end (E2E) network slicing orchestrator that effectively allocates resources across the three network domains. Its performance has been validated through real hardware implementation, demonstrating reliable resource isolation per slice [19].

The business model related to network slicing comprises several integral components and entities vital to its functioning. These include the network slicing instance (NSI), network slicing subnet instance (NSSI), logical network, network subslice (NSS), network slicing template (NST), network segment, Network Function Virtualization (NFV), Software-Defined Networking (SDN), network slicing manager, communication service manager, resource slice, network slicing provider, network slicing terminal, network slicing tenant, network slicing repository, slice border control, slice selection function, infrastructure owner, infrastructure slice, infrastructure slice provider, and infrastructure slice tenant [20][21]. Each of these elements has specific roles and capabilities in managing the life cycle of a network slice. As an example, the NSI denotes a group of comprehensive logical networks that provide various services tailored to specific requirements, encompassing multiple sub-slice instances. Conversely, the NSSI represents the localized logical network within a network slice, which can be shared among multiple NSIs. Logical networks are virtual instances of network functions established on a single physical network.

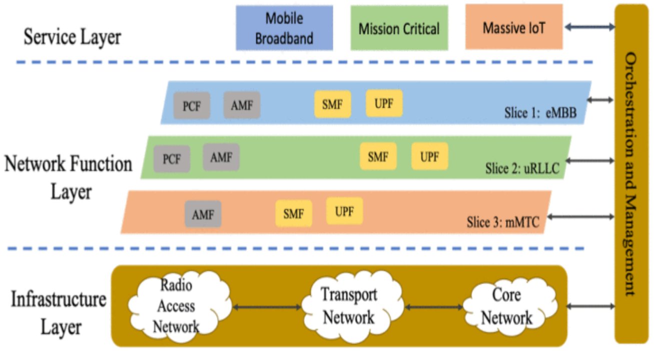


Fig. 1. Network slicing for 5g networks, adopted from [13].

The network slicing manager plays a pivotal role in overseeing the complete lifecycle of each slice or sub-slice by performing various management functions. These include the communication service management function (CSMF), network slice management function (NSMF), and network slice subnet management function (NSSMF). CSMFs are responsible for managing, communicating, and updating the requirements of the slice to support service requests through the communication service manager. NSMFs handle the management of NSIs based on the notifications received from CSMFs, while NSSMFs manage the NSSIs according to the requirements specified by NSMFs. The network slicing requirements encompass various aspects such as network type, network capacity, quality of service (QoS), latency, security level, device count, and throughput. The resource slice refers to the combination of physical and virtual resources necessary for the functioning of network slices. The network slicing provider is the entity responsible for owning the physical infrastructure where multiple slices are created, whereas the network slicing tenant refers to the users of the NSI who deliver specific services requested by customers. The infrastructure owner denotes the entity that owns the physical infrastructure, On the other hand, the infrastructure slice provider is the entity that owns the infrastructure and leases it to host a variety of services via network slicing. The infrastructure slice tenants are the users of the infrastructure slice itself [13].

IV. Security in Network Slicing

Ensuring security is of utmost importance during the implementation of network slicing, which enables the support of diverse services with varying security requirements [22]. The utilization of network slicing in multi-domain infrastructures, serving multiple customers, can introduce complex security challenges. This is especially evident when resources are shared among slices that adhere to distinct security policies established by various verticals and operators. To effectively address security issues both within and among slices, it is crucial to consider security coordination and protocols during the resource allocation and design stages. Neglecting these aspects may result in the emergence of new and advanced security vulnerabilities in 5G systems and beyond [23]. Each slice is created with isolation constraints in order to prevent the propagation of attack impacts across slices and allow for independent security solutions [24]. It is essential to ensure adherence to fundamental security principles such as confidentiality, authentication, availability, integrity, and authorization within each slice [25]. Confidentiality safeguards against unauthorized data disclosure, authentication verifies the identities of parties involved in interactions, availability ensures the accessibility of slices and applications, integrity guarantees that slice owners maintain control over functionalities and configurations, and authorization determines the permissible capabilities for each network element. Availability refers to the ability of the system to meet the requirements of service level agreements by ensuring that slices and applications can be accessed as needed, while Network Slice Manager (NSM) and Network Functions (NFs) remain consistently accessible. On the other hand, integrity ensures that only slice owners possess the authority to modify or replace the functionality and configuration of their respective slices [26]. Authorization defines the permitted functionalities for each network element, where slice owners are responsible for managing and controlling their respective slices, end-users engage exclusively with authorized slices, infrastructure providers oversee the network slice management (NSM), NSM governs the network slice instances (NSIs) and network functions, and network functions exercise control over resources. These elements consist of slice owners, end-users, service providers, infrastructure providers, network slice management (NSM), and Network Function Virtualization (NFV). The security of network slicing requires each slice and its owners to independently fulfill these requirements in order to mitigate the potential exploitation of network slicing features by attackers, which could lead to system failures [13].

A. Security Issues Introduced by Network Slicing

Network Slicing is distinguished by its crucial attribute of isolation, which has a direct impact on the dependability of the slicing solution. Attaining a high level of isolation is essential to achieve optimal outcomes. Merely supporting a single slice in a slicing system would essentially replicate a conventional non-sliced network, which is already extensively studied. Hence, the coexistence of multiple slices becomes a necessary prerequisite for network slicing, wherein these slices share the same underlying infrastructure. The ability to coexist without interference hinges on establishing the minimum requirements for each slice. By meeting these requirements, interference can be avoided, thereby ensuring effective isolation. Ensuring security in network slicing involves precisely defining the boundaries of interference for each slice, specifying the minimum requirements, and enforcing compliance with these requirements [27].

The significance of identifying isolation characteristics, implementing an abstraction layer for achieving end-to-end isolation at an appropriate level, and establishing appropriate security policies was emphasized in a study [28]. This survey highlighted the lack of a standardized description for isolation capabilities that can be employed for automated deployment. Hence, it is crucial to define the desired initial level of isolation, especially concerning security, and devise dynamic isolation mechanisms capable of enforcing the required level of isolation for each specific service. To tackle the security concerns associated with network slicing in the context of 5G, several organizations, such as the Next Generation Mobile Networks (NGMN), have released guidelines [29]. These recommendations assist in identifying potential security risks in the general packet core. Additionally, technology guidelines, such as ETSI's recommendations for Network Function Virtualization (NFV), provide further guidance in addressing security concerns. [30], have been taken into account. ETSI's guidelines encompass security considerations throughout the lifecycle of virtual network functions.

B. More Security Challenges to Network Slicing

Network slicing implementation in the telecommunications industry presents challenges, particularly concerning security and the deployment of the radio access network (RAN). These challenges have been extensively discussed by researchers like Kotulski et al. [32]. Security concerns arise with the introduction of network slicing. One significant risk is the potential for denial-of-service (DoS) attacks and resource depletion [32], which can disrupt network availability and performance. Other threats include monitoring, traffic injection, and impersonation attacks [27], jeopardizing data integrity and network efficiency. Moreover, the use of diverse systems from different vendors in network slicing creates a security vulnerability due to the absence of standardized security measures [32]. Exploiting vulnerabilities in these systems, attackers can successfully target network slices. Additionally, specific weaknesses in the isolation and protection mechanisms designed for network slices may enable unauthorized access to resources and sensitive data. To address these security threats and vulnerabilities, robust security measures are necessary. These can involve implementing intrusion detection and prevention systems, enforcing access control policies, and establishing comprehensive security monitoring. Furthermore, integrating security considerations into the design of network slicing systems and protocols ensures a secure-by-default approach. Standardized security practices enhance overall network slicing security and promote interoperability among different vendor systems [31] [32]. The physical realization of the RAN poses challenges in implementing network slicing. Ensuring resource, traffic, and user isolation within radio network elements is a significant obstacle. One suggested approach involves the utilization of millimeter waves for covering small cells, leveraging the cell size to achieve isolation. Nevertheless, current technologies like cognitive radio and non-orthogonal multiple access fall short in delivering the necessary levels of isolation and slicing required to meet the desired standards. Thus, further research and development are essential to overcome the physical implementation challenges of network slicing in the RAN [31].

The challenges pertaining to security and access management in network slicing originate from the varying security and privacy requirements [27] of different network slices in 5G networks. Network slicing is dependent on the implementation of software-defined networking (SDN) and network function virtualization (NFV). [33], which virtualize network functions as software components. However, ensuring secure inter-slice access and end-to-end network slicing security becomes more complex in 5G due to its comprehensive slicing approach. Securing the management of network slicing encompasses various tasks, including establishing secure access between the radio access network and core network resources, ensuring secure connections between user equipment (UE) and network slice instances, and effectively managing shared resources among different network slices. Neglecting these challenges can lead to security risks, including compromised controllers or orchestrators, isolation failures, insider threats, compromised NFV instances, and unauthorized data access [34][35]. The centralized slice manager also presents security considerations, including the security of network slice templates, concerns regarding unauthorized access, and issues related to trust [36]. Additionally, in multi-domain infrastructures or multi-tenant environments, network slicing may encounter additional security and privacy challenges. To mitigate these concerns, it is essential to adhere to established security principles, including but not limited to confidentiality, integrity, authenticity, availability, and authorization [37][38]. In the context of network function virtualization (NFV), maintaining confidentiality is vital to mitigate potential threats. For example, a compromised slice manager can lead to unauthorized monitoring of traffic through both northbound and southbound interfaces, potentially exposing sensitive slice configurations. Similarly, vulnerabilities in the application programming interface (API) used during configuration can enable malicious actors to interfere with slice installation, configuration, or activation. Thus, ensuring the confidentiality of inter-slice communications is crucial for establishing a secure environment [39][40][41]. Proactive measures like implementing secure communication protocols, enforcing access controls, and conducting regular security policy reviews are necessary to mitigate these threats effectively.

V. Methodology

This section provides an overview of the methodology employed to conduct a Systematic Literature Review (SLR) on the subject of Machine Learning-based secure 5G network slicing, following the recommended guidelines outlined in [42]. The process of formulating research questions is discussed, along with the motivating factors behind these questions. Various data sources were used to select relevant articles, and a specific search strategy was employed to obtain articles in the domain. Inclusion and exclusion criteria were applied to select articles for review. In order to present a comprehensive overview of the current state-of-the-art in Machine Learning-based 5G network slicing, Table I illustrates the research questions along with their respective motivations.

TABLE I. Research Questions and Motivation

	Question	Motivation
RQ1	What are the most advanced machine learning techniques currently employed for ensuring the security of 5G network slicing?	It aims to identify and analyze existing machine learning-based techniques for securing 5G network slicing to develop more effective and efficient security solutions.
RQ2	What are the main obstacles and constraints faced by these techniques in terms of accuracy, scalability, and efficiency?	It aims to identify the potential issues that may arise when implementing machine learning-based techniques for securing 5G network slicing, and to develop strategies to overcome these challenges and limitations.
RQ3	What are the potential solutions to address these challenges and limitations?	It aims to identify and propose possible solutions/recommendations to overcome the challenges and limitations of machine learning-based techniques for securing 5G network slicing.
RQ4	What are the future research directions and opportunities in this area?	It aims to explore the integration of explainable AI and federated learning, investigate the transferability of models across different network architectures and scenarios, and design new evaluation metrics and methodologies for machine learning-based solutions.

A. Data Sources

Table II displays the reputable publishers such as IEEE, Science Direct, Springer, ACM Digital Library, Wiley, Sage, MDPI and Google Scholar, from which the articles were selected for review.

TABLE II. Database Sources

Publisher	URL
IEEE	https://www.ieee.org
Science Direct	https://www.sciencedirect.com
Springer	https://link.springer.com
ACM Digital Library	https://www.acm.org
Wiley	https://onlinelibrary.wiley.com
Sage	https://journals.sagepub.com
MDPI	https://www.mdpi.com
Google Scholar	https://scholar.google.com/

B. Search Strategy

Due to limited research in this area initially, articles considered for review in this study were limited to those published from the year 2018 onwards. The initial stage in constructing the search query involved identifying appropriate keywords aligned with the theme and the research questions put forth. Primary keywords including "network slicing," "5G," "security," and "machine learning" were identified, and logical operators such as "AND" and "OR" were used to link these keywords appropriately. After conducting several tests, the researchers arrived at a search string that produced a sufficient number of related research articles. The keywords used in the search string are listed in Table III.

TABLE III. SEARCH STRING

String	Batch1	Batch2	Batch3	Batch4
String1	Network Slicing	5G	Security	Machine Learning
String2	Network-Slicing			Machine-Learning
String3				ML

C. Article Selection Process

The methodology employed for selecting articles commenced with formulating research questions that guided the creation of a search query for article retrieval. Only articles published in the English language were taken into account. The PRISMA flow diagram was utilized to visualize the article selection process [43] to ensure a systematic and transparent approach to article selection, as illustrated in Fig. 2. Once the pertinent literature was identified, a comprehensive review of load balancing techniques in software-defined networks was performed. The review process concluded with a meticulous categorization of the load balancing techniques to ensure thoroughness. Many articles were excluded during the screening process due to their titles not meeting the inclusion criteria or abstracts not being relevant for the survey.

Search String: (([Batch1, String1] OR [Batch1, String2]) AND ([Batch2, String1]) AND ([Batch3, String1]) AND ([Batch4, String1] OR [Batch4, String2] OR [Batch4, String3])).

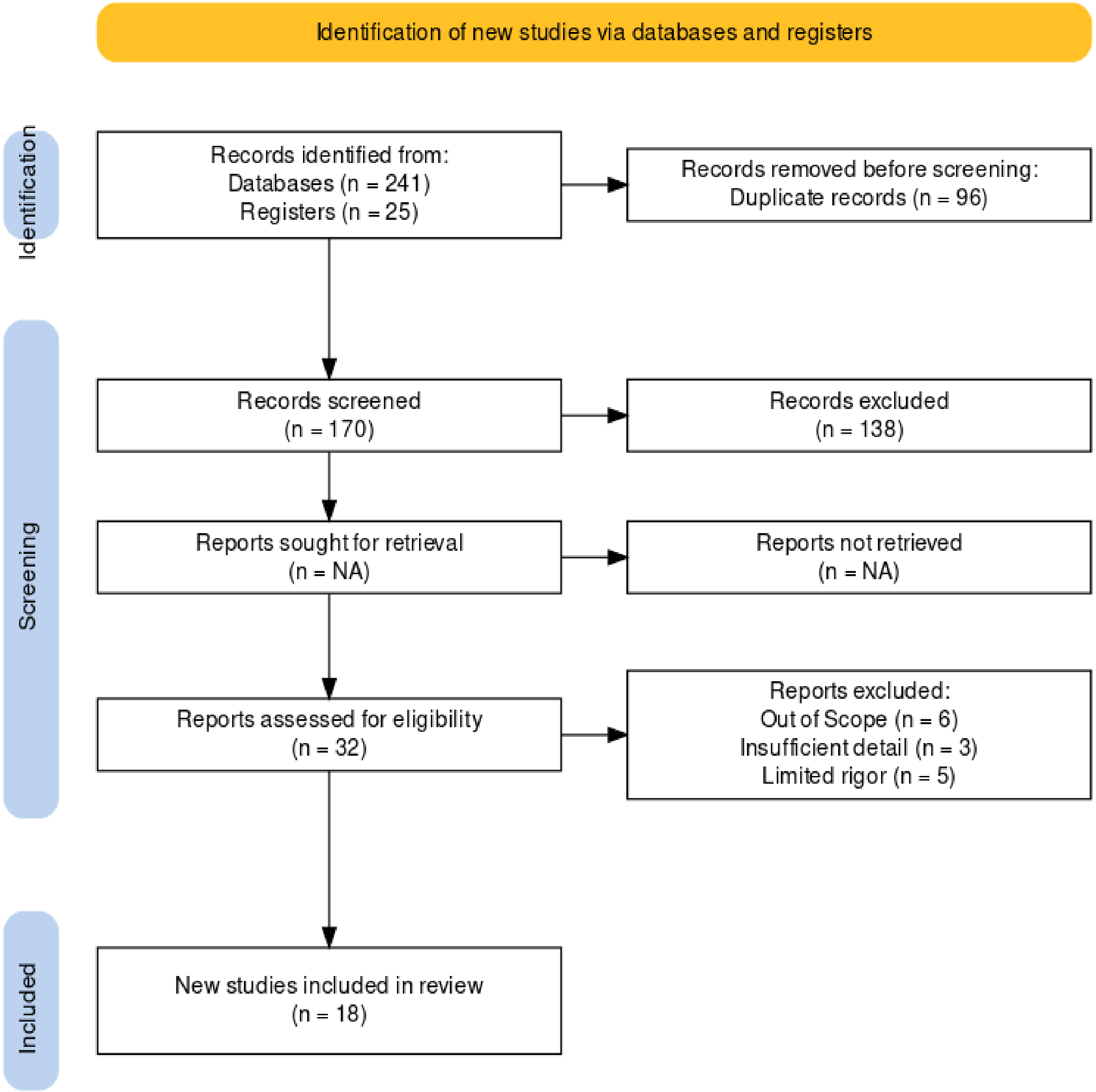


Fig. 2. PRISMA flow diagram.

D. Data Extraction

The research methodology utilized in this study encompassed an extensive exploration to collect relevant literature pertaining to the subject of secure 5G network slicing utilizing machine learning. The primary objective was to identify and select articles that provide valuable insights into this field of study. The research process consisted of several distinct phases, including a targeted search, application of inclusion and exclusion criteria, and comprehensive evaluation of selected articles. The initial step in the research process involved performing a targeted search using specific keywords. This search was conducted across reputable publishers and covered the period between 2018 and 2023. Fig. 3 shows the year-wise number of articles that were published by these publishers. The list of sources used to perform this survey is provided in Table II. The aim was to gather a comprehensive collection of articles related to the subject matter. As a result of this search, a total of 241 articles were identified and retrieved.

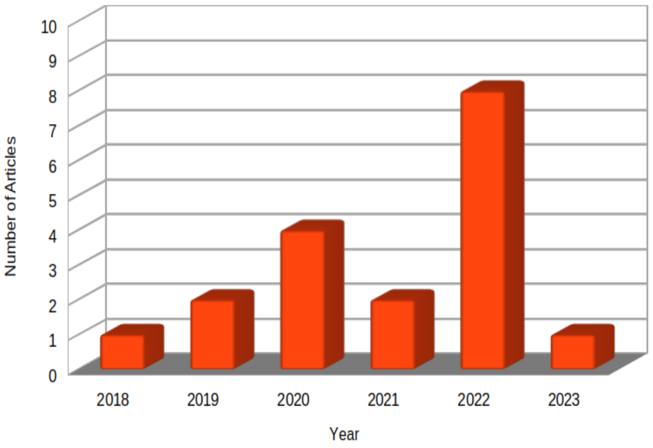


Fig. 3. Year-wise categorization of articles.

E. Inclusion and Exclusion Criteria

To refine the selection and focus on significant research, inclusion and exclusion criteria were established. These criteria were specifically designed to guarantee the inclusion of only articles that directly relate to the research topic. Table IV presents a comprehensive breakdown of the inclusion and exclusion criteria implemented during this process. By applying these criteria, the initial pool of 241 articles was reduced to 170, thereby eliminating articles that did not meet the predefined criteria. The subsequent phase of the research involved a detailed examination of the remaining 170 articles. During this phase, the titles and abstracts of the articles underwent a thorough examination to further refine the selection process. The purpose of this review was to identify articles that aligned closely with the research focus on machine learning-based secure 5G network slicing. As a result of this review, 32 articles were selected for further evaluation.

The selected 32 articles underwent a comprehensive evaluation based on their content to ensure a close match with the objectives and scope of the current research. This evaluation involved an in-depth analysis of the full texts of the articles, including an examination of the methodology, findings, and discussions presented. The aim of this evaluation was to identify articles that provide valuable insights and contribute directly to the research topic. Following this rigorous evaluation process, a final set of 18 articles was identified as the most relevant to the research topic. The selection of these 18 essential research articles was conducted meticulously, taking into account the alignment between the titles, abstracts, and comprehensive content of the articles. The selected articles were recognized as valuable sources of insights into machine learning-based secure 5G network slicing, and their inclusion in the study was deemed essential for the progression of the current research endeavor.

TABLE IV. INCLUSION AND EXCLUSION CRITERIA

Inclusion	Exclusion
The study focuses on Machine Learning-based secure 5G network slicing	The study that focuses on areas other than Machine Learning-based secure 5G network slicing
Only the articles written in English language are considered	Articles written in non-English language are not considered
Articles published by the publishers listed in Table II	Unpublished articles and those that are not peer-reviewed, are not considered
Articles published in well-reputed and high impact factor journals are considered	White papers, editorials, keynote speeches, and articles from predatory journals are not considered

VI. DISCUSSION

In this systematic literature review, we investigated the current state of research on machine learning-based solutions for secure 5G network slicing. Our review identified a total of 18 relevant articles that met our inclusion criteria. From our analysis of these articles, several key themes emerged. Firstly, machine learning techniques are widely used for various security-related tasks in network slicing, including anomaly detection, intrusion detection, and malware detection. Secondly, while there is a broad consensus on the potential benefits of machine learning-based solutions for securing 5G network slicing, there is also a lack of standardization and interoperability among different solutions. Finally, the use of machine learning in network slicing introduces several challenges related to data privacy, explainability, and scalability. In this discussion section, we will elaborate on these themes while answering the research questions and provide recommendations for future research in this area.

Q1. What is the most advanced machine learning techniques currently employed for ensuring the security of 5G network slicing?

Network slicing is a vital component of 5G and future generation networks since it enables the creation of customized logical networks with diverse functionalities, dependability, and security properties. Scholars have proposed multiple types of Machine Learning (ML) and Deep Learning (DL) approaches to increase the security and reliability of network slicing. Table V summarizes the state-of-the-art ML/DL algorithms used in the scientific literature for secure 5G network slicing. This discussion examines the strategies for secure 5G network slicing presented in the selected articles. Fig. 4 illustrates the performance of various machine learning algorithms concerning the security of 5G network slicing. Each algorithm is evaluated based on three key criteria: accuracy, scalability, and efficiency. Accuracy reflects the algorithm's ability to provide precise and reliable security measures. Scalability measures its capacity to handle an increasing number of network slices efficiently, accounting for complex dependencies. Efficiency assesses how well the algorithm utilizes resources during security processes. The chart clearly distinguishes the strengths and weaknesses of different algorithms. For instance, algorithms like RL (Reinforcement Learning) and DQN (Deep Q-Network) demonstrate high scores in all three categories, making them robust choices for secure 5G network slicing. On the other hand, algorithms like LSTM (Long Short-Term Memory) and DBN (Deep Belief Networks) show comparatively lower scores, suggesting room for improvement in their application.

TABLE V. STATE-OF-THE-ART ML TECHNIQUES, CHALLENGES AND POTENTIAL SOLUTIONS

Ref.	Algorithm	Challenges			Potential Solutions
		Accuracy	Scalability	Efficiency	
[45][49][51][52][61]	CNN (Convolutional Neural Networks)	Lack of accuracy in network slice orchestration and optimization and deep learning-based DDoS attack detection for network slicing.	Scalability challenges in securing multiple network slices simultaneously	Inefficient resource utilization during attack mitigation	Integrating reinforcement learning techniques with CNNs can address accuracy, scalability, and resource utilization challenges in secure 5G network slicing, including network slice orchestration, DDoS attack detection, and simultaneous protection of multiple network slices.
[44][55][56][62]	RL (Reinforcement Learning)	Limited accuracy in intelligent resource allocation	Difficulty in scaling up network slices due to complex dependencies	Inefficient resource utilization during IP shuffling processes	Multi-agent RL for accurate, scalable, and efficient secure 5G network slicing for intelligent resource allocation, complex dependencies, and IP shuffling processes.
[46][52][54][66]	SVM (Support Vector Machine)	Challenges in achieving accurate end-to-end security in network slices	Scalability challenges in managing security across numerous network slices	Inefficient utilization of security resources	Distributed and scalable SVM frameworks can address accuracy and scalability challenges in securing network slices, optimizing resource utilization in 5G network slicing.
[50][57][60][68]	DQN (Deep Q-Network)	Accuracy challenges in context-aware authentication handover for secure network slicing	Scalability concerns in managing authentication handover across numerous network slices	Inefficient resource allocation for authentication handover processes	Application of advanced deep reinforcement learning techniques, such as hierarchical or multi-agent DQN, to improve accuracy, scalability, and resource allocation for context-aware authentication handover in secure 5G network slicing.
[45][49][69]	LSTM (Long Short-Term Memory)	Lack of accuracy in network slice orchestration and optimization	Scalability issues in managing a large number of network slices	Inefficient resource allocation and utilization	Advanced LSTM-based deep learning with RL/attention improves accuracy, scalability, and resource allocation in secure 5G network slice orchestration.
[58][59]	FL (Federated Learning)	Limited local data in network slices hampers model performance and accuracy.	Limited computational resources in network slices hinder scalability, impacting Federated Learning.	Sub-optimal resource allocation and usage can impact system efficiency and performance	Improve model performance with data augmentation and transfer learning, optimize resource allocation for scalability, and employ efficient scheduling techniques to enhance system efficiency in secure 5G network slicing.
[48]	K-means clustering, Naive Bayes classifier	Accuracy issues in automated machine learning for network slice automation	Scalability concerns in managing and orchestrating a large number of network slices	Inefficient utilization of automated processes	Improving accuracy of automated machine learning models; Scalable orchestration frameworks; Efficient utilization of automated processes through optimization
[53]	GAN (Generative Adversarial Networks)	Accuracy challenges in adversarial machine learning for flooding attacks in network slicing	Scalability issues in detecting and mitigating flooding attacks across multiple network slices	Inefficient resource utilization during attack mitigation	Improving accuracy of adversarial machine learning models; Scalable flooding attack detection mechanisms; Efficient utilization of resources during attack mitigation
[54]	DBN (Deep Belief Networks)	Accuracy issues due to complex network slices, limited labeled data, and difficulty capturing intricate patterns.	Scalability issues in managing numerous slices efficiently and handling increased computational and communication overhead.	Inefficient resource utilization, communication overhead, and synchronization processes.	Data augmentation, transfer learning, efficient parallelization, optimized resource allocation, and adaptive strategies to improve accuracy, scalability, and resource utilization.

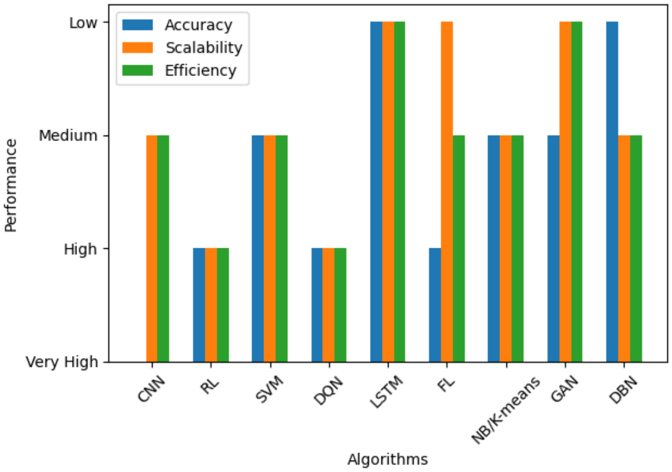


Fig. 4. Algorithm performance for secure 5G network slicing.

Jiang et al. [44] proposed a "Intelligence Slicing" framework combining network slicing and AI for improved intelligence, security, and flexibility. To maximize the efficiency of network slicing and to ensure end-to-end security, the framework employs AI techniques such as reinforcement learning, transfer learning, and deep learning. Deep neural networks are used in [45] to predict resource requirements for different slices, and a clustering technique is used to group analogous slices. The goal is to eliminate wastage of resources while maintaining network stability. Liu et al. [46] suggested a learning-assisted secure end-to-end network slicing technique for cyber-physical systems (CPS). For efficient resource allocation among similar CPSs, ML approaches such as k-means clustering and k-nearest neighbors are used. To maintain slice confidentiality and integrity, a secure key exchange mechanism is used. Sedjelmaci [47] presented a cooperative attack detection system based on AI, applying machine learning techniques such as random forest and decision trees to identify any type of malicious activity and differentiate between legitimate and malicious traffic. This method improves network slicing security by detecting and mitigating threats more effectively. Kafle et al. [48] developed an ML-based network slicing automation strategy based on a multi-agent reinforcement learning technique. To provide end-to-end security, the approach learns from previous slicing events to make accurate slicing decisions while optimizing resource allocation and restricting the attack surface. Secure5G is a DL-based architecture for secure network slicing that was introduced in [49]. It makes use of deep neural networks to estimate slice resource requirements and efficient resource allocation. The framework consists of a security module that detects and mitigates network attacks using a deep belief network. A reinforcement learning-based technique for attacking and defending 5G radio access network slicing is proposed in [50]. It learns optimal attack and defense techniques through deep reinforcement learning, enabling network administrators to identify and mitigate vulnerabilities. The DeepSecure technique introduced in [51] analyzes traffic patterns and detects distributed denial-of-service (DDOS) attacks on 5G network slicing, using convolutional neural networks. It enhances reliability and security by effectively detecting and mitigating such attacks. A hybrid DL-based approach for wireless network slicing is proposed in [52]. This approach combines deep neural networks for network traffic prediction and a reinforcement learning algorithm for resource allocation optimization. It aims to improve the reliability and accuracy of network slicing while ensuring end-to-end security. In [53], an adversarial ML approach is proposed to detect and mitigate flooding attacks on 5G radio access network slicing. It enhances security by effectively identifying and mitigating such attacks. Benzaid et al. [54] propose an AI-based autonomic security management architecture for secure network slicing in B5G networks, focusing on effective security management leveraging AI techniques. Another approach proposed in [56] introduces a learning augmented optimization approach to safeguard network slicing in 5G. It presents a mathematical model and optimization algorithm that utilizes machine learning to adapt to network environment changes, minimizing network cost while ensuring secure slices meeting quality of service requirements. [57] presents SliceBlock, a secure network slicing scheme utilizing a DAG-blockchain to authenticate handover requests between network slices in edge-assisted SDN/NFV-6G environments. It integrates DAG-blockchain technology with SDN/NFV and edge computing, providing a secure and reliable network slicing service. Federated learning, proposed in [59], aims to improve security in network slicing by aggregating training data from multiple slices to detect anomalies and security threats across the entire network. Simulation experiments demonstrate its effectiveness in enhancing security threat detection accuracy. Lastly, [60] proposes a reinforcement learning approach for attacking and defending NextG radio access network slicing. It utilizes multi-agent reinforcement learning to train agents capable of launching attacks and defending against them in network slicing scenarios.

Q2. What are the main obstacles and constraints faced by these techniques in terms of accuracy, scalability, and efficiency?

Network slicing has become a pivotal concept within 5G networks, enabling operators to divide their networks into virtualized networks that are customized to meet the distinct needs of various applications and services. Integrating machine learning (ML) and deep learning (DL) techniques presents a promising avenue for enhancing the security of network slicing. This integration facilitates the detection and prevention of security threats in real-time, offering an effective approach to safeguarding network slicing. However, leveraging ML and DL in secure network slicing poses various challenges, including the need for labeled data, complexity of models, and ensuring data confidentiality and privacy. Subsequent sections will delve into a comprehensive examination of these challenges, providing a detailed analysis, and presenting potential solutions to address them.

Jiang et al. [44] propose a comprehensive framework named "Intelligence slicing" that combines artificial intelligence (AI) with 5G networks. However, they do not discuss the limitations in terms of accuracy and scalability associated with their framework. Thantharate et al. [45] propose "DeepSlice," an approach based on deep learning (DL) for achieving efficient and dependable network slicing in

5G networks. The authors recognize the challenges associated with handling large volumes of training data and the need for substantial computational resources, which could potentially impede scalability and efficiency. Liu et al. [46] introduce the concept of "Learning-assisted secure end-to-end network slicing" for cyber-physical systems. They emphasize the issue of explainability in machine learning-based security approaches, underscoring its potential impact on trust and the acceptance of such methods. Sedjelmaci [47] puts forward a cooperative attack detection system based on artificial intelligence (AI) for 5G networks. The authors highlight the drawbacks of conventional rule-based systems and emphasize the potential advantages of AI-based systems, but they do not explicitly address the specific limitations associated with such approaches. Kafle et al. [48] suggest the automation of 5G network slicing through the utilization of machine learning techniques. The authors recognize the challenges brought about by the complexity of the network and the requirement for substantial amounts of training data. Thantharate et al. [49] introduce "Secure5G," a DL framework for secure network slicing in 5G and future networks. However, they recognize the challenge of developing DL models capable of detecting and defending against new and unknown attacks. The research in [50] presents a reinforcement learning approach for attacking and defending 5G radio access network slicing, but does not delve into the accuracy or scalability limitations of the approach. Kuadey et al. [51] introduce "DeepSecure," which is a deep learning-based approach for detecting distributed denial-of-service (DDoS) attacks in 5G network slicing. The authors acknowledge the difficulty of developing accurate models while minimizing false positives. The study presented in [52] proposes a hybrid deep learning approach for achieving high accuracy and reliability in wireless network slicing within the context of 5G. However, the study does not delve into the limitations of this approach. Shi and Sagduyu the approach proposed in [53] is based on adversarial machine learning and aims to address flooding attacks on 5G radio access network slicing. However, the study does not explicitly discuss the challenge of developing robust models that are resilient to adversarial attacks. Benzaid et al, [54] introduces an AI-based autonomic and scalable security management architecture for secure network slicing in B5G. However, they do not specifically discuss the limitations related to developing models capable of handling the dynamic and heterogeneous nature of B5G networks.. Yoon et al. [55] presents a technique called "Moving target defense for in-vehicle software-defined networking" that utilizes IP shuffling in network slicing, combined with multi-agent deep reinforcement learning. They address the challenges associated with accurate attack detection and mitigation in dynamic and mobile environments. Cheng et al. [56] suggests a method called "Safeguard network slicing in 5G" that employs a learning-augmented optimization approach. They acknowledge the challenge of managing the complexity and dynamics inherent in 5G networks. Abdulqadder and Zhou [57] present "SliceBlock," a solution that combines context-aware authentication handover and secure network slicing using DAG-blockchain in edge-assisted SDN/NFV-6G environments. However, the study does not delve into the specific limitations associated with their proposed models. In their paper [58], Bandara et al. introduces a federated learning platform for network slicing that incorporates blockchain and zero trust security mechanisms. However, the proposed platform encounters challenges in terms of ensuring high accuracy and scalability. Similarly, Wijethilaka and Liyanage [59] suggest a federated learning approach to enhance security in network slicing. However, the approach faces challenges in terms of scalability and efficiency. On the other hand, Shi et al. [60] utilize reinforcement learning for attacking and defending NextG radio access network slicing. However, their proposed approach encounters limitations in terms of scalability and accuracy. Lastly, Chowdhury et al. [61] introduce AUTODEEPSLICE, a data-driven technique for network slicing that employs automatic deep learning. However, they face challenges in achieving accurate results due to the inherent complexity of network slicing.

Q3. What are the potential solutions to address these challenges and limitations?

1) Solutions presented in the literature: One possible solution to tackle the challenges and limitations discussed in the papers involves developing and implementing intelligent slicing frameworks that integrate artificial intelligence (AI) into 5G networks [44]. This strategy capitalizes on machine learning algorithms to optimize and automate network slicing processes [48], enhancing the reliability of network slicing [52], and fortifying the security of 5G networks against cyber-attacks [46][49][51]. An alternative strategy is to utilize reinforcement learning to protect against flooding attacks in 5G radio access network slicing [53][60][63][64][65]. Furthermore, a context-aware authentication handover and secure network slicing that incorporates a DAG-blockchain in edge-assisted SDN/NFV-6G environments can enhance the security of network slicing [57]. Federated learning can also be deployed to bolster the security of network slicing [59]. By incorporating the Skunk-A Blockchain and Zero Trust Security Enabled Federated Learning Platform, the security of 5G/6G network slicing is significantly enhanced. [58]. Finally, data-driven network slicing techniques utilizing automatic deep learning can be employed to enhance network slicing in 5G networks [61].

2) Recommended solutions: The integration of Convolutional Neural Networks (CNNs) with reinforcement learning techniques offers a promising solution to tackle challenges related to accuracy, scalability, and resource utilization in secure 5G network slicing effectively. This integration can be applied to various aspects, including network slice orchestration, DDoS attack detection, and simultaneous protection of multiple network slices.

a) Markov Decision Process (MDP): Consider the problem of network slice orchestration as a Markov Decision Process (MDP) [63], defined by the tuple (S, A, P, R, γ). Here, S is the state space, A is the action space, P represents the transition probabilities, R is the reward function, and γ is the discount factor. The objective is to find an optimal policy π that maximizes the expected cumulative reward:

$$\pi^* = argmax_\pi E_\pi[\sum_{t=0}^{\infty} \quad \gamma^t R(s_t, a_t)] \tag{1}$$

Where, π^* represents the optimal policy that provides the best actions a_t in each state s_t, taking into account the long-term cumulative reward. The MDP formulation considers network slice orchestration as a sequential decision-making problem. By optimizing the policy π, it ensures that actions taken in each state lead to the maximum cumulative reward, which is vital for efficient network resource allocation.

b) Bayesian reinforcement learning: To improve DDoS attack detection, we can model it as a Bayesian Reinforcement Learning problem [64], where we need to infer the optimal policy π given a sequence of observations O and actions A. We seek to maximize the posterior probability of the policy given the data:

$$P(\pi|O,A) \propto P(O|\pi,A)P(\pi|A) \qquad (2)$$

Where, $P(\pi|O,A)$ is the posterior policy probability, $P(O|\pi,A)$ is the likelihood of observations given the policy, and $P(\pi|A)$ is the prior policy probability. The Bayesian RL formulation models DDoS attack detection as a probabilistic learning problem. It estimates the posterior policy π given observed data, enabling improved detection accuracy through probabilistic reasoning.

c) Multi-objective optimization: For simultaneous protection of multiple network slices, we can formulate it as a Multi-Objective Optimization problem [65]. Let $F = [f_1(x), f_2(x), \ldots, f_k(x)]$ represent a vector of k objective functions, and $X = [x_1, x_2, \ldots, x_n]$ denote the vector of decision variables. Additionally, $G = [g_1(x), g_2(x), \ldots, g_m(x)]$ represents a vector of m inequality constraints. The goal is to find a vector of decision variables X^* that optimizes multiple objectives while satisfying constraints:

$$X^* = arg\ min_x F(X) = [f_1(X), f_2(X), \ldots, f_k(X)] \qquad (3)$$

Subject to:

$$g_i(X) \leq 0, for\ i = 1,2,\ldots,m$$

The multi-objective optimization approach provides a rigorous framework to balance multiple objectives while considering constraints. It ensures that resources are efficiently allocated to protect multiple network slices.

d) Multi-agent reinforcement learning: In the context of secure 5G network slicing, multi-agent reinforcement learning [66] demonstrates potential for addressing accuracy, scalability, and resource utilization challenges. Specifically, it can enable intelligent resource allocation by considering complex dependencies and facilitating IP shuffling processes. We performed an experiment using python to demonstrate a custom reinforcement learning environment, termed NetworkSlicingEnv, designed to emulate a 5G network slicing scenario, where an agent makes resource allocation decisions with the goal of optimizing the allocation of resources based on dependencies among actions. The experiment utilizes the Proximal Policy Optimization (PPO) algorithm, a cutting-edge reinforcement learning method. The agent's training process includes learning to optimize resource allocation while considering dependencies among actions, contributing to the achievement of optimal resource utilization and scalability in 5G network slicing.

We define the ‘NetworkSlicingEnv’ environment and then train a PPO model to optimize resource allocation within the environment. After training, we evaluate the model's performance by calculating the mean reward over a specified number of episodes, denoted by ‘n_episodes’. This mean reward serves as an indicator of the model's resource allocation capabilities in the secure 5G network slicing scenario. The algorithm presented below depicts the scenario:

```
Algorithm:
Inputs:
    'n_episodes': Number of episodes for evaluation
Outputs:
    'mean_reward': Mean reward over the evaluation episodes
Initialize:
  1. Define the environment class 'NetworkSlicingEnv':
     a. Initialize the action space with 10 discrete resource allocation options.
     b. Initialize the observation space with a 5-dimensional state observation.
Procedure 'TrainModel' (Total Timesteps: 10,000):
  1. Create a multi-agent environment instance 'env'.
  2. Define a PPO model with a multi-layer perceptron policy ('MlpPolicy') for policy optimization.
  3. Train the model with the 'learn' method using 'total_timesteps=10,000'.
Procedure 'EvaluateModel' (Number of Episodes: 'n_episodes'):
  1. Initialize 'mean_reward' to 0.
  2. For each episode in the range 'n_episodes', do the following:
     a. Reset the environment ('env') and obtain the initial state.
     b. For each time step:
        i. Predict an action using the trained model ('model') based on the current state.
        ii. Simulate the environment's response and receive the reward.
        iii. Accumulate the reward.
        iv. Check if the episode is done. If done, break the loop.
  3. Calculate the 'mean_reward' as the sum of rewards over 'n_episodes' divided by 'n_episodes'.
Output 'mean_reward'.
End
```

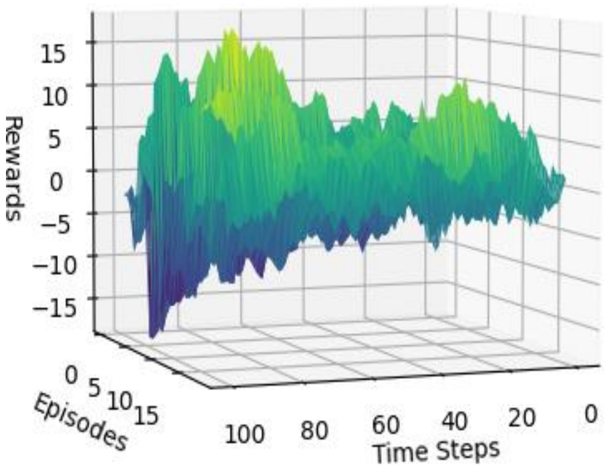


Fig. 5. Multi-Agent reinforcement learning training rewards over time.

The experimental results reveal the agent's learning progress, represented through a 3D plot showcasing the cumulative reward over training steps as shown in Fig. 5. The positive rewards indicate that the agent is learning to make resource allocation decisions that result in outcomes deemed favorable. This implies that the agent is successfully adapting its behavior over time to maximize the cumulative rewards. The learning dynamics suggest that the agent is gradually improving its understanding of the environment and making better choices in allocating resources. As the training progresses, the agent becomes more proficient in resource allocation. The positive rewards show that the agent is making decisions that lead to efficient resource utilization. This can be seen as an indication of improved performance in managing network resources for 5G network slicing. The 3D plot visualizes how the agent's rewards change over the course of training. Gradually, the agent should learn to make resource allocation decisions that result in higher cumulative rewards, which signifies an improved understanding of the environment and more effective decision-making. The mean reward, calculated at the end of training, provides a summary of the agent's performance. A higher mean reward suggests that the agent has improved its decision-making skills regarding resource allocation, which is a critical aspect of network slicing in 5G environments. The positive rewards demonstrate that the agent's learning dynamics are effective in improving its performance in resource allocation decision-making. The agent gradually learns to allocate resources more efficiently, resulting in higher cumulative rewards and, by extension, better performance for secure 5G network slicing.

e) Support Vector Machine (SVM): For securing network slices in a scalable manner, distributed and scalable Support Vector Machine (SVM) [67] frameworks can be employed. These frameworks contribute to improving both accuracy and scalability while optimizing resource utilization within the 5G network slicing environment. In the experiment conducted, a synthetic dataset was generated for binary classification, simulating a scenario in which SVM-based classification is employed. The code utilizes the Scikit-Learn library to create a two-dimensional dataset with 1000 samples. The algorithm presented below depicts the scenario:

```
Algorithm:
Input:
        • Training data: $(X_{train}, y_{train})$
        • Testing data: $(X_{test}, y_{test})$
Initialization:
    a. Initialize SVM model: $SVM_{model}$.
Data Preparation:
    a. Split the data into training and testing sets:
        • $X_{train}, y_{train}$ for training.
        • $X_{test}, y_{test}$ for testing.
Accuracy Optimization:
    a. Train the SVM model on the training data:
        • $SVM_{model}.fit(X_{train}, y_{train})$.
    b. Make predictions on the test data:
        • $y_{pred} = SVM_{model}.predict(X_{test})$.
    c. Calculate accuracy ($A$) using the ground truth and predictions:
        • $A = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$.
Scalability Optimization:
    a. In SVM, the margin is automatically maximized during training to improve scalability ($S$).
Resource Utilization:
    a. SVMs automatically select a subset of support vectors for decision function, which reduces resource utilization ($R$).
Output:
    a. Return:
        • Trained SVM model: $SVM_{model}$.
        • Accuracy ($A$).
        • Scalability ($S$).
        • Resource Utilization ($R$).
End
```

The data is divided into training and testing sets to evaluate the SVM model's performance. The SVM model employs a linear kernel for classification and is trained on the training data. The decision boundary is plotted along with the data points, with a color scale indicating the accuracy level of 0.88 in this case. The visualization demonstrates how SVM, as a scalable framework, can effectively classify data while maintaining a high level of accuracy as depicted in Fig. 6. This technique proves to be an optimal solution for ensuring network slice security and efficient resource allocation within the complex 5G environment, making this experiment a valuable illustration of SVM's capabilities in a network slicing context.

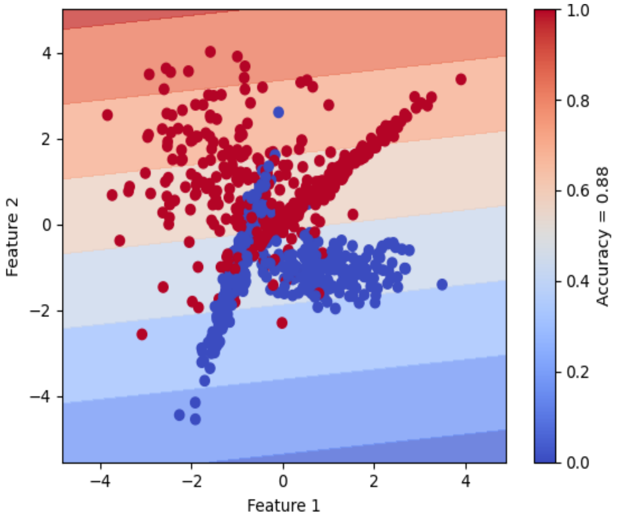


Fig. 6. Secure network slice classification: SVM decision boundary with 88% accuracy.

f) Deep learning: By leveraging sophisticated deep reinforcement learning methods like hierarchical or Multi-Agent Deep Q-Networks (MADQN) [68], it is possible to improve accuracy, scalability, and resource allocation in the context of context-aware authentication handover within secure 5G network slicing. We used TensorFlow library to demonstrate the concept of rewards over a series of episodes, environments, which is a critical aspect of evaluating the performance of Multi-Agent Deep Q-Networks (MADQN) in context-aware authentication handover within 5G network slicing environments. These rewards are indicative of the efficiency and learning progress of a group of agents working collaboratively to make decisions. The rewards represent various factors, including the success of authentication handover processes, the minimization of resource utilization, and the overall optimization of network slicing. By examining the rewards over episodes, we gain valuable insights into how well the MADQN is adapting and learning within this environment as shown in Fig. 7. Consistent, positive rewards indicate that the MADQN is effectively enhancing the authentication handover process and resource allocation, contributing to scalability and the overall security of 5G network slicing.

```
Algorithm:
Input:
    ● State size (‘state_size’)
    ● Action size (‘action_size’)
    ● Number of agents (‘num_agents’)
    ● Total training episodes (‘episodes’)
Initialization:
    ● Initialize ‘state_size’, ‘action_size’, ‘num_agents’, and
      ‘episodes’.
    ● Initialize the multi-agent environment (‘env’) with
      given parameters.
Agent Creation:
    ● For each agent (‘i = 1’ to ‘num_agents’):
        a. Create a DQNAgent (‘agent[i]’) with
           ‘state_size’ and ‘action_size’.
        b. Build a neural network model for ‘agent[i]’ with
           three dense layers.
Training Loop:
    ● For each episode (‘e = 1’ to ‘episodes’):
        a. Generate random initial states for each agent and
           store in states.
        b. For each step in the episode:
            i. Query each agent for actions based on their current
               state.
            ii. Implement the logic to take actions and receive new
                states, rewards, and 'done' flags.
            iii. Update each agent's Q-network based on their
                 experiences.
            iv. If all agents are done (all 'done' flags are True), exit
                the loop.
Output:
    ● Trained MADQN agents.
End
```

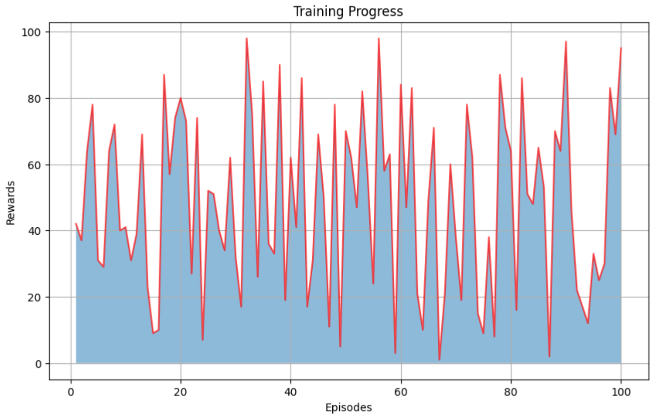


Fig. 7. Multi-Agent Deep Q-Networks (MADQN) episode rewards over time.

Furthermore, the utilization of advanced Long Short-Term Memory (LSTM)-based deep learning architectures [69], integrated with reinforcement learning or attention mechanisms, shows promise in improving accuracy, scalability, and resource allocation for secure 5G network slice orchestration.

g) Other potential solutions: To further improve the performance of secure 5G network slicing, data augmentation and transfer learning techniques can enhance model performance, while optimized resource allocation ensures scalability. Efficient scheduling techniques contribute to enhancing system efficiency within the secure 5G network slicing context. Moreover, potential solutions for improving the accuracy of automated machine learning models, scalable orchestration frameworks, and efficient utilization of automated processes through optimization are essential considerations. Additionally, addressing the accuracy of adversarial machine learning models, deploying scalable flooding attack detection mechanisms, and efficiently utilizing resources during attack mitigation are crucial aspects to enhance the security and performance of secure 5G network slicing. Overall, employing techniques such as data augmentation, transfer learning, efficient parallelization, optimized resource allocation, and adaptive strategies holds promise for enhancing accuracy, scalability, and resource utilization in the secure 5G network slicing domain.

Fig. 8. Word cloud visualization of potential solutions for secure 5G network slicing

The word cloud depicted in Fig. 8 is a visual representation of potential solutions for enhancing secure 5G network slicing, as discussed in the literature. The above word cloud was generated using python. This visualization method succinctly conveys the key strategies and concepts described in the literature by assigning word sizes based on their frequency of occurrence. Larger and bolder words indicate the prominence and importance of specific terms within the text. Notably, terms such as "security," "reinforcement learning," "slicing frameworks," "resource utilization," and "scalability" are the most prominent in the word cloud. These words represent the primary solutions put forward in the literature for addressing the challenges and limitations related to secure 5G network slicing. The word cloud effectively distills the essence of the text, providing an at-a-glance overview of the critical ideas and strategies to enhance accuracy, scalability, and resource utilization in the secure 5G network slicing domain. Scientists and researchers can quickly grasp the core themes and solutions discussed in the paper, making it a valuable addition to the scientific discourse for understanding complex topics with enhanced clarity.

Q4. What are the future research directions and opportunities in this area?

1) Future directions as discussed in the literature: The papers surveyed in this study present several potential future research directions in the domain of secure 5G network slicing. One promising avenue is the integration of reinforcement learning techniques with convolutional neural networks (CNNs) to address challenges related to accuracy, scalability, and resource utilization. This approach has shown promise in enhancing network slice orchestration, enabling DDoS attack detection, and facilitating simultaneous protection of multiple network slices [44]. Another area of interest is multi-agent reinforcement learning, which can be explored to tackle accuracy, scalability, and resource utilization challenges in secure 5G network slicing. By leveraging intelligent resource allocation and effectively managing complex dependencies, multi-agent reinforcement learning offers the potential for more efficient and effective network slicing [45].

Furthermore, the literature suggests the exploration of distributed and scalable frameworks such as support vector machines (SVM) to address accuracy and scalability challenges in securing network slices and optimizing resource utilization in 5G network slicing [46]. Additionally, advanced deep reinforcement learning techniques, including hierarchical or multi-agent deep Q-networks (DQN), can be applied to improve accuracy, scalability, and resource allocation in context-aware authentication handover for secure 5G network slicing [52][68]. The use of advanced LSTM-based deep learning architectures, coupled with reinforcement learning or attention mechanisms, also holds promise for enhancing accuracy, scalability, and resource allocation in secure 5G network slice orchestration [69].

2) Recommended future directions

a) Ensemble learning techniques: Ensemble learning (stacking or boosting for accuracy and robustness enhancement) involves combining multiple individual models to create a stronger and more accurate predictive model. Stacking and boosting are two common ensemble techniques. Stacking combines the predictions of multiple models using another model, often referred to as a meta-learner. Boosting, on the other hand, assigns more weight to instances that are misclassified by the previous models in the ensemble, thereby focusing on improving the accuracy of these instances. The application of ensemble techniques in secure 5G network slicing aims to enhance prediction accuracy and overall robustness by leveraging the strengths of multiple models.

b) Integration of explainable AI: Explainable AI techniques (for transparency and trust) aim to provide understandable and interpretable explanations for the decisions made by machine learning models. In the context of secure 5G network slicing, incorporating explainable AI methods allows users to understand the rationale behind decisions made by the network slicing system. This transparency enhances trust and accountability, particularly in critical applications where the ability to explain decisions is essential. By enabling stakeholders to comprehend how decisions are reached, explainable AI techniques contribute to increased confidence and acceptance of the system's outcomes.

c) Federated learning: Federated learning (for data privacy and security) is a decentralized machine learning approach that enables model training across multiple devices or nodes while keeping data localized. This approach prioritizes data privacy and security by not requiring data to be centralized for training. In the context of secure 5G network slicing, federated learning offers a solution to challenges related to sharing sensitive data across different slices. It allows models to be trained collaboratively without sharing raw data, thus ensuring privacy while still improving the performance of network slicing algorithms.

d) Edge computing: Utilization of edge computing and edge intelligence for localized decision-making involves processing data closer to the data source, reducing communication overhead and latency. Incorporating edge intelligence in network slicing enables localized decision-making at the edge nodes, which can lead to quicker responses and efficient resource utilization. By processing data and making decisions closer to where it is generated, edge computing optimizes the performance of network slicing and enhances its responsiveness.

e) Energy efficiency: Energy efficiency through algorithmic development and resource allocation is a critical concern in any network infrastructure. In the context of secure 5G network slicing, developing energy-efficient algorithms and resource allocation strategies is essential to minimize power consumption while maintaining optimal performance. By optimizing the allocation of resources based on workload and demand, energy-efficient network slicing can contribute to reducing operational costs and improving overall sustainability.

f) Generative adversarial networks: Generative Adversarial Networks (GANs) are a type of machine learning model used in tasks such as image generation and data synthesis. In the context of secure 5G network slicing, GANs can be employed to generate realistic adversarial examples that simulate potential attacks on the network slicing algorithms. By testing the robustness of these algorithms against such simulated attacks, GANs can help identify vulnerabilities and weaknesses in the system's security measures, thereby enhancing its overall security posture.

VII. Conclusion

This literature survey presented a comprehensive overview of the research conducted in the domain of secure 5G network slicing. The reviewed papers highlighted the challenges related to accuracy, scalability, and resource utilization in network slicing and proposed various techniques to address these issues. The experimental results show that the integration of reinforcement learning techniques with CNNs, multi-agent reinforcement learning, and distributed SVM frameworks emerged as potential solutions to improve accuracy and scalability in network slicing. Advanced deep reinforcement learning architectures, such as hierarchical or multi-agent DQN, and LSTM-based models with reinforcement learning or attention mechanisms were identified as effective approaches for enhancing accuracy, scalability, and resource allocation in network slice orchestration. Furthermore, data augmentation, transfer learning, efficient parallelization, optimized resource allocation, and adaptive strategies were suggested as methods to improve model performance, scalability, and resource utilization. The surveyed literature also shed light on the importance of securing network slices against cyber threats, with studies exploring the detection and mitigation of DDoS attacks, cooperative attacks, and adversarial machine learning models. Additionally, considerations for privacy, explainability, edge computing, energy efficiency, and the application of emerging techniques like Ensemble Learning, Federated learning, GANs, etc. were highlighted as potential future research directions. By exploring these avenues, this review has laid a foundation for researchers to contribute to the advancement of secure 5G network slicing, addressing the challenges and ensuring the reliability, efficiency, and security of future network infrastructures.

This review explored the significant advancements and challenges in the integration of 5G networks with security, network slicing, and machine learning techniques. The findings reveal that 5G networks offer immense potential for delivering high accuracy, scalability, and efficiency in various applications. Network slicing emerges as a crucial mechanism for resource allocation and management in 5G networks, enabling efficient utilization of network resources for different services. Moreover, machine learning and deep learning algorithms demonstrate promising capabilities in enhancing network performance and security by enabling intelligent decision-making and anomaly detection. However, the integration of these technologies also presents notable challenges, such as ensuring robust security measures, optimizing network slicing algorithms, and addressing scalability issues. Further research and development efforts are required to overcome these challenges and fully exploit the potential of 5G networks with security, network slicing, and machine learning for various domains and applications.

Acknowledgment

This project was funded by Deanship of Scientific Research, Northern Border University for their financial support under grant no. (NR-NBU-2023-12-2125). The authors, therefore, acknowledge with thanks DSR technical and financial support.

References

[1] X. Li et al., “Network slicing for 5G: Challenges and opportunities,” IEEE Internet Computing, pp. 1–1, 2018. doi:10.1109/mic.2018.326150452.

[2] H. Yu, H. Lee, and H. Jeon, “What is 5G? emerging 5G Mobile Services and network requirements,” Sustainability, vol. 9, no. 10, p. 1848, 2017. doi:10.3390/su9101848.

[3] R. Dangi et al., “ML-based 5G network slicing security: A comprehensive survey,” Future Internet, vol. 14, no. 4, p. 116, 2022. doi:10.3390/fi14040116.

[4] T. Chhabra, “5G in India: The journey is about to begin - ET telecom,” ETTelecom.com, https://telecom.economictimes.indiatimes.com/news/5g-in-india-the-journey-is-about-to-begin/81671088 (accessed Feb. 12, 2023).

[5] T. GreyB, “5G companies: 12 players are leading the research,” GreyB, https://www.greyb.com/blog/5g-companies/ (accessed Feb. 9, 2023).

[6] G. Narcisi, “These are the 5G trends to watch in 2021,” CRN, https://www.crn.com/news/networking/these-are-the-5g-trends-to-watch-in-2021 (accessed Feb. 10, 2023).

[7] “What is 5g: Everything you need to know About 5G: 5G FAQ: Qualcomm,” Wireless Technology & Innovation, https://www.qualcomm.com/5g/what-is-5g# (accessed Feb. 9, 2023).

[8] N. Brittain, “18 5G projects providing a vision for the future,” 5Gradar, https://www.5gradar.com/features/5g-projects-that-will-blow-your-mind (accessed Feb. 14, 2023).

[9] H. Zhang et al., “Network slicing based 5G and future mobile networks: Mobility, Resource Management, and challenges,” IEEE Communications Magazine, vol. 55, no. 8, pp. 138–145, 2017. doi:10.1109/mcom.2017.1600940.

[10] C. Campolo, A. Molinaro, A. Iera, and F. Menichella, “5G network slicing for vehicle-to-everything services,” IEEE Wireless Communications, vol. 24, no. 6, pp. 38–45, 2017. doi:10.1109/mwc.2017.1600408.

[11] G. Werélius, What we know: A look at current 5G market trends - ericsson, https://www.ericsson.com/en/blog/2020/10/what-we-know-a-look-at-current-5g-market-trends (accessed Feb. 9, 2023).

[12] R. F. Olimid and G. Nencioni, “5G network slicing: A security overview,” IEEE Access, vol. 8, pp. 99999–100009, 2020. doi:10.1109/access.2020.2997702.

[13] F. Salahdine, Q. Liu, and T. Han, “Towards secure and intelligent network slicing for 5G networks,” IEEE Open Journal of the Computer Society, vol. 3, pp. 23–38, 2022. doi:10.1109/ojcs.2022.3161933.

[14] E. P. Neto, F. S. Silva, L. M. Schneider, A. V. Neto, and R. Immich, “Seamless Mano of multi-vendor SDN controllers architectures and future challenges,” Comput. Netw., vol. 167, no. 106984, p. 106984, 2020.

[15] S. Chaabnia and A. Meddeb, “Slicing aware QoS/QoE in Software Defined Smart Home Network,” NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, 2018. doi:10.1109/noms.2018.8406195.

[16] P. K. Chartsias et al., “SDN/NFV-based end to end network slicing for 5G multi-tenant networks,” in 2017 European Conference on Networks and Communications (EuCNC), 2017.

[17] C. Bektas, S. Monhof, F. Kurtz, and C. Wietfeld, “Towards 5G: An empirical evaluation of software-defined end-to-end network slicing,” in 2018 IEEE Globecom Workshops (GC Wkshps), 2018.

[18] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, “5G network slicing using SDN and NFV: A survey of taxonomy, domains,” IEEE Access, vol. 8, pp. 29525–29537, 2020.

[19] X. Li, R. Ni, J. Chen, Y. Lyu, Z. Rong, and R. Du, “End-to-end network slicing in radio access network, transport network and core networkacross Federated Multi-domains,” Computer Networks, vol. 186, p. 107752, 2021. doi:10.1016/j.comnet.2020.107752.

[20] S. D’Oro, F. Restuccia, A. Talamonti, and T. Melodia, “The slice is served: Enforcing radio access network slicing in virtualized 5G systems,” in IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, 2019.

[21] A. Kaloxylos, “A survey and an analysis of network slicing in 5G networks,” IEEE Commun. Stand. Mag., vol. 2, no. 1, pp. 60–65, 2018.

[22] F. Salahdine and N. Kaabouch, “Security threats, detection, and countermeasures for physical layer in cognitive radio networks: A survey,” Phys. Commun., vol. 39, no. 101001, p. 101001, 2020.

[23] M. A. Habibi, B. Han, and H. D. Schotten, “Network slicing in 5G mobile communication architecture, profit modeling, and challenges,” arXiv [cs.NI], 2017.

[24] N. Alliance, “5G security recommendations package# 2: Network slicing,” NGMN, pp. 1–12, 2016.

[25] S. Sharma et al., “Secure authentication protocol for 5G enabled IoT network,” in 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2018.

[26] N. Alliance, “Description of network slicing concept,” NGMN 5G P, vol. 1, no. 1, pp. 1–11, 2016.

[27] V. A. Cunha et al., “Network slicing security: Challenges and directions,” Internet Technol. Lett., vol. 2, no. 5, p. e125, 2019.

[28] Z. Kotulski et al., “Towards constructive approach to end-to-end slice isolation in 5G networks,” EURASIP J. Inf. Secur., vol. 2018, no. 1, 2018.

[29] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network slicing in 5G: Survey and challenges,” IEEE Commun. Mag., vol. 55, no. 5, pp. 94–100, 2017.

[30] GROUP SPECIFICATION, “ETSI GS NFV-SEC 001 V1.1.1 (2014-10),” Etsi.org. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/nfv-sec/001_099/001/01.01.01_60/gs_nfv-sec001v010101p.pdf. [Accessed: 10-April-2023].

[31] A. Mathew, “Network slicing in 5G and the security concerns,” in 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020.

[32] Z. Kotulski et al., “On end-to-end approach for slice isolation in 5G networks. Fundamental challenges,” in Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, 2017.

[33] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, “Network slicing and softwarization: A survey on principles, enabling technologies, and solutions,” IEEE Commun. Surv. Tutor., vol. 20, no. 3, pp. 2429–2453, 2018.

[34] F. Reynaud et al., “Attacks against network functions virtualization and software-defined networking: state-of-the-art,” in Proceedings of Workshop on Security in Virtualized Networks, Sec-Virtnet, 2016.

[35] V. N. Sathi, M. Srinivasan, P. K. Thiruvasagam, and S. R. M. Chebiyyam, “A novel protocol for securing network slice component association and slice isolation in 5G networks,” in Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2018.

[36] “Security aspects of network capabilities exposure in 5G,” in Final deliverable (approved-P Public), 2018.

[37] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, “A survey on security and privacy of 5G technologies: Potential solutions, recent advancements, and future directions,” IEEE Commun. Surv. Tutor., vol. 22, no. 1, pp. 196–248, 2020.

[38] S. Zhou, L. Wu, and C. Jin, “A privacy-based SLA violation detection model for the security of cloud computing,” China Commun., vol. 14, no. 9, pp. 155–165, 2017.

[39] S. Y.-T. Fan C-I, “Cross-network-slice authentication scheme for the 5th generation mobile communication system,” IEEE Transactions on Network and Service Management, 2021.

[40] “Security,” 5G Second Phase Explained, pp. 235–258, 2021. doi:10.1002/9781119645566.ch7.

[41] S. Bao, Y. Liang, and H. Xu, “Blockchain for network slicing in 5G and beyond: Survey and challenges,” J. Commun. Inf. Netw., vol. 7, no. 4, pp. 349–359, 2022.

[42] S. K. Boell and D. Cecez-Kecmanovic, ‘‘On being ‘systematic’ in literature reviews,’’ in Formulating Research Methods for Information Systems. London, U.K.: Palgrave Macmillan, 2015, pp. 48–78.

[43] M. D. J. Peters, C. M. Godfrey, H. Khalil, P. McInerney, D. Parker, and C. B. Soares, ‘‘Guidance for conducting systematic scoping reviews,’’ Int. J. Evidence-Based Healthcare, vol. 13, no. 3, pp. 141–146, Sep. 2015.

[44] W. Jiang, S. D. Anton, and H. Dieter Schotten, “Intelligence slicing: A unified framework to integrate artificial intelligence into 5G networks,” in 2019 12th IFIP Wireless and Mobile Networking Conference (WMNC), 2019.

[45] A. Thantharate, R. Paropkari, V. Walunj, and C. Beard, “DeepSlice: A deep learning approach towards an efficient and reliable network slicing in 5G networks,” in 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2019.

[46] Q. Liu, T. Han, and N. Ansari, “Learning-assisted secure end-to-end network slicing for cyber-physical systems,” IEEE Netw., vol. 34, no. 3, pp. 37–43, 2020.

[47] H. Sedjelmaci, “Cooperative attacks detection based on artificial intelligence system for 5G networks,” Comput. Electr. Eng., vol. 91, no. 107045, p. 107045, 2021.

[48] V. P. Kafle, Y. Fukushima, P. Martinez-Julia, and T. Miyazawa, “Consideration on automation of 5G network slicing with machine learning,” in 2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K), 2018.

[49] A. Thantharate, R. Paropkari, V. Walunj, C. Beard, and P. Kankariya, “Secure5G: A deep learning framework towards a secure network

slicing in 5G and beyond," in 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 2020.

[50] How to attack and defend 5G radio access network slicing with reinforcement learning. arXiv 2021.

[51] N. A. E. Kuadey, G. T. Maale, T. Kwantwi, G. Sun, and G. Liu, "DeepSecure: Detection of distributed denial of service attacks on 5G network slicing—deep learning approach," IEEE Wirel. Commun. Lett., vol. 11, no. 3, pp. 488–492, 2022.

[52] Highly accurate and reliable wireless network slicing in 5th generation networks: a hybrid deep learning approach, Journal of Network and Systems Management: Springer, 2022.

[53] Y. Shi and Y. E. Sagduyu, "Adversarial machine learning for flooding attacks on 5G radio access network slicing," in 2021 IEEE International Conference on Communications Workshops (ICC Workshops), 2021.

[54] C. Benzaid, T. Taleb, and J. Song, "AI-based autonomic and scalable security management architecture for secure network slicing in B5G," IEEE Netw., vol. 36, no. 6, pp. 165–174, 2022.

[55] S. Yoon et al., "Moving target defense for in-vehicle software-defined networking: IP shuffling in network slicing with multiagent deep reinforcement learning," in Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II, 2020.

[56] X. Cheng, Y. Wu, G. Min, A. Y. Zomaya, and X. Fang, "Safeguard network slicing in 5G: A learning augmented optimization approach," IEEE J. Sel. Areas Commun., vol. 38, no. 7, pp. 1600–1613, 2020.

[57] I. H. Abdulqadder and S. Zhou, "SliceBlock: Context-aware authentication handover and secure network slicing using DAG-blockchain in edge-assisted SDN/NFV-6G environment," IEEE Internet Things J., vol. 9, no. 18, pp. 18079–18097, 2022.

[58] E. Bandara, X. Liang, S. Shetty, R. Mukkamala, A. Rahman, and N. W. Keong, "Skunk-A Blockchain and Zero Trust Security Enabled Federated Learning Platform for 5G/6G Network Slicing," in 2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), IEEE, 2022, pp. 109–117.

[59] S. Wijethilaka and M. Liyanage, "A federated learning approach for improving security in network slicing," in GLOBECOM 2022 - 2022 IEEE Global Communications Conference, 2022.

[60] Y. Shi, Y. E. Sagduyu, T. Erpek, and M. C. Gursoy, "How to attack and defend NextG radio access network slicing with reinforcement learning," IEEE Open J. Veh. Technol., vol. 4, pp. 181–192, 2023.

[61] D. Chowdhury, R. Das, R. Rana, A. D. Dwivedi, P. Chatterjee, and R. R. Mukkamala, "AUTODEEPSLICE: A Data Driven Network Slicing Technique of 5G network using Automatic Deep Learning," in 2022 IEEE Globecom Workshops (GC Wkshps), 2022.

[62] J. Wang and J. Liu, "Secure and reliable slicing in 5G and beyond vehicular networks," IEEE Wirel. Commun., vol. 29, no. 1, pp. 126–133, 2022.

[63] L. Tang, Q. Tan, Y. Shi, C. Wang, and Q. Chen, "Adaptive virtual resource allocation in 5G network slicing using constrained Markov decision process," IEEE Access, vol. 6, pp. 61184–61195, 2018.

[64] V. Sciancalepore, X. Costa-Perez, and A. Banchs, "RL-NSB: Reinforcement learning-based 5G network slice broker," IEEE ACM Trans. Netw., vol. 27, no. 4, pp. 1543–1557, 2019.

[65] G. Zhou, L. Zhao, G. Zheng, Z. Xie, S. Song, and K. C. Chen, "Joint Multi-objective Optimization for Radio Access Network Slicing Using Multi-agent Deep Reinforcement Learning," IEEE Transactions on Vehicular Technology, 2023.

[66] Y. Kim and H. Lim, "Multi-agent reinforcement learning-based resource management for end-to-end network slicing," IEEE Access, vol. 9, pp. 56178–56190, 2021.

[67] O. A. Latif, M. Amer, and A. Kwasinski, "Classification of network slicing requests using support vector machine," in 2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA), 2022.

[68] P. Tam, S. Math, A. Lee, and S. Kim, "Multi-agent deep Q-networks for efficient edge federated learning communications in software-defined IoT," Comput. Mater. Contin., vol. 71, no. 2, pp. 3319–3335, 2022.

[69] R. Li, C. Wang, Z. Zhao, R. Guo, and H. Zhang, "The LSTM-based advantage actor-critic learning for resource management in network slicing with user mobility," IEEE Commun. Lett., vol. 24, no. 9, pp. 2005–2009, 2020.

Comparison of the Application of Weighted Cosine Similarity and Minkowski Distance Similarity Methods in Stroke Diagnostic Systems

Joko Purwadi[1], Rosa Delima[2], Argo Wibowo[3], Angelina Rumuy[4]
Informatics Department, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia[1, 2, 4]
Information System, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia[3]

***Abstract*—Stroke is a critical medical condition requiring prompt intervention due to its multifaceted symptoms and causes influenced by various factors, including psychological aspects and the patient's lifestyle or daily habits that impact risk factors. The recovery process involves consistent medical care and lifestyle adjustments tailored to the individual case. Expert Systems, a scientific field focused on studying and developing diagnostic systems, can employ the Case-based Reasoning method to identify the type of stroke based on similarities with prior patient cases, considering specific causes and symptoms. This study utilizes the Weighted Cosine, Jaccard Coefficient, and Minkowski Distance methods to assess the similarity of stroke cases. The evaluation is based on input data such as patient causes or symptoms and risk factors from medical records. The analysis of case similarity and solutions involves applying the Weighted Cosine, Jaccard Coefficient, and Minkowski Distance methods, with a defined threshold value. The highest similarity values from previous patient cases are selected for each method. The test outcomes suggest that employing the Minkowski Distance method with a threshold value of 75 and an r value of three or four yields the highest levels of accuracy, recall, and precision. The Minkowski Distance achieves an accuracy and recall rate of more than 88 percent with 100 percent precision.**

Keywords—Expert system; stroke; case-based reasoning; Minkowski Distance; jaccard coefficient; weighted cosine; threshold; accuracy; diagnosis

I. INTRODUCTION

Stroke is an emergency disease that must be treated immediately to minimize brain damage due to lack of oxygen and nutrients. Stroke can cause paralysis and death for patients. According to [1], stroke is an non-communicable condition arising from blockage, constriction, or hemorrhage within the brain's blood vessels, resulting in a diminished blood flow to the brain.

The development of information technology specifically for intelligent systems or artificial intelligence (AI) has had a positive impact on progress in the field of medicine and health. One of the applications of intelligent systems to support disease diagnosis is the development of an Expert System. According to reference [2][3], the expert system is a system based on knowledge that utilizes expert knowledge to address a particular issue. The essential components required for constructing the expert system include a knowledge base, inference engine, working memory, and user. There has been a number of systems developed in the medical field including [4] [5] [6] [7] [8] [9] [10] [11].

The utilization of techniques in creating the expert system is highly varied, with one example being the Case-Based Reasoning (CBR) approach. CBR involves retrieving cases from past occurrences, subsequently reusing and adapting them in new situations. [12] [13].

In the academic realm, expert systems play a crucial role in the learning process, especially in the field of stroke-related studies. The medical histories of individuals who have experienced strokes can be employed as a point of reference when diagnosing new cases, utilizing the Case-Based Reasoning (CBR) method. In this CBR process, successfully resolved issues are archived for potential use in future problem-solving scenarios. Conversely, if a problem persists without resolution, the case is identified and stored to prevent similar errors in the future [14]. The CBR method comprises four key stages of problem-solving: retrieve, reuse, revise, and retain.

Retrieve is taking back cases most similar or relevant to the new case [15]. Meanwhile, reuse is the process of reusing information and knowledge from old cases as solutions for new cases. The old case, which has a similarity value above the threshold value and which has the highest value, is reused as a solution to solve the new case. Revise is the revision process that involves reviewing and revising the proposed solution. In this process, information is re-evaluated to address problems that arise in new cases. After that, the system will generate a solution for the new problem [16], and retaining is the process of storing new cases that have been successfully resolved and have found solutions to the database so that they can be reused as solutions for new cases in the future.

There has been many fields that apply CBR to solve problems. In the geographical field, Dou, et al. [17] conducted research to detect landslides using CBR. Tempola and colleagues [18] conducted a study investigating the use of Case-Based Reasoning (CBR) to assess the qualification of students for scholarship awards.

In the CBR system, the calculation of case similarity at the retrieve phase becomes a very important part. This calculation is the basis for determining the level of document similarity. There are several similarity calculation metrics, such as Minkowski Distance similarity and weighted cosine similarity.

The Minkowski Distance represents a generalized version of the Euclidean Distance and Manhattan Distance approaches. [19]. The main difference lies in the value of r, which is a power constant in the *Minkowski Distance* method. Meanwhile, *Weighted Cosine Similarity* calculates the similarity between two objects based on the size of the cosine angle [20]. The primary objective of this study is to assess and compare the effectiveness of the Minkowski Distance Similarity and Weighted Cosine Similarity methods in achieving the highest accuracy for the stroke diagnosis system.

Another research was conducted by Adawiyah [21] regarding the use of the *Minkowski Distance* for the system to detect premature baby birth. A system accuracy testing was carried out using 20 test data, seven data with a normal diagnosis and 13 data with a premature diagnosis. From the test, there was two data obtained whose results are not appropriate because the value is below the *threshold*, which is ≤ 60%. The accuracy of the system is 90% in detecting premature births.

In 2022, Mubarak, et al. conducted a research on CBR for the diagnosis of malaria using the *Minkowski Distance Similarity* method where testing was carried out with 25 test data and 58 training data which showed a system accuracy value of 92% with a *threshold* of 80% [22].

A research on case-based reasoning to diagnose malnutrition in children aged 0 - 5 years by applying the Cosine Similarity method [23] was conducted by Soinbala, et al. in 2019. The system's accuracy and validity were evaluated through testing with 40 new cases, comparing the system's diagnostic results with those provided by experts. The test outcomes reveal an 80% accuracy rate when employing an 80% threshold.

Zainuddin and colleagues conducted a study in 2016 [24], concentrating on Case-Based Reasoning (CBR) for diagnosing strokes. They employed the K-Nearest Neighbor algorithm with an 80% threshold value. The research, based on 15 test cases, reported a system accuracy of 93.3%, consistent with expert diagnoses. In a separate study, Warman and team [25] investigated the use of expert systems in identifying diseases in rice plants. They utilized CBR with the K-Nearest Neighbor method for distance calculation. The evaluation of system sensitivity and accuracy involved 52 test data points with a threshold value of 70%. The findings indicated a system sensitivity of 100% and an accuracy rate of 82.69%.

This study represents a continuation of the research conducted by Nelson et al. in 2018. In their investigation, Nelson et al. developed an expert system for diagnosing strokes using a Case-Based Reasoning (CBR) approach. The system employs the Jaccard Coefficient method for calculating case similarity. The research utilized the Siriraj score as a distinguishing factor between ischemic and hemorrhagic types of strokes, incorporating dense indexing for enhanced efficiency [26]. The system underwent testing with 45 cases as test data and utilized 135 cases as a case base. The findings revealed that a threshold value of 0.7 resulted in superior sensitivity and accuracy compared to threshold values of 0.8, 0.9, and 1. The system demonstrated a sensitivity level of 89.88% and accuracy of 81.67% with indexing and 84.44% without indexing. Further research was carried out using the same dataset with the Minkowski Distance similarity calculation method [27]. The research results show that Minkowski Distance provides a better accuracy rate of 88.89% compared to the Jaccard Coefficient method.

The research carried out is a continuation of research [26] [27]. The focus of this research is to compare the level of similarity test between the Minkowski Distance Similarity and Weighted Cosine Similarity methods in the diagnosis of stroke patients. This research wants to find out whether Weighted Cosine Similarity can increase the accuracy of the system in diagnosing stroke. This research contributes to increasing the effectiveness of expert systems in diagnosing stroke.

II. Method

The process of developing the system involves several stages, commencing with needs analysis, followed by system design, program code implementation, and culminating in system testing, as depicted in Fig. 1.

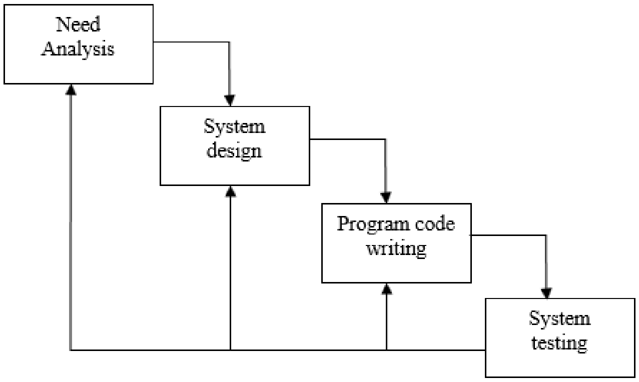


Fig. 1. The system development stages.

A. Data

The data used in this study was sourced from Nelson et al.'s 2018 research [26], specifically the study titled "Case-Based Reasoning for Stroke Diseases Diagnosis." It encompasses medical records extracted from patients who had experienced strokes and were treated at Dr. Soetarto DKT Hospital in Yogyakarta during the period of 2015-2016. The data were categorized into four types of stroke based on both the cause and anatomical pathology, namely embolic stroke, thrombotic stroke, subarachnoid hemorrhage stroke, and intracerebral hemorrhage stroke. We have limitations in terms of test data and future work has the opportunity to carry out better and more complete tests.

B. System Planning

The expert system is developed utilizing the Case-Based Reasoning (CBR) method, integrating the Minkowski Distance similarity method to evaluate similarities between newly entered cases and existing ones. Users input information in the form of the patient's personal data, symptoms, and risk factors. Subsequently, the system calculates local and global similarity values between the newly entered case data (user-provided data) and the cases stored in the case base. The case exhibiting the highest similarity, surpassing a predetermined threshold, is applied as the solution for new cases. In instances where the similarity value falls below the threshold, the case is retained in the case base for expert review. Conversely, if the similarity exceeds the threshold, the system generates an output indicating the type of stroke affecting the patient. The use case diagram for stroke diagnosis in the expert system is depicted in Fig. 2.

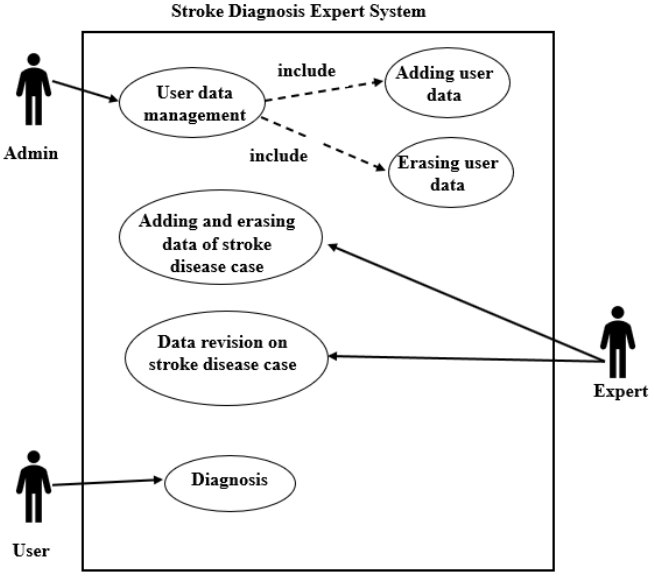

Stroke Diagnosis Expert System
Admin
User data management
include
Adding user data
include
Erasing user data
Adding and erasing data of stroke disease case
Data revision on stroke disease case
Expert
Diagnosis
User

Fig. 2. The expert system usecase diagram for stroke diagnosis.

The goal of similarity measurement is to evaluate how closely two objects resemble each other. The determination of the similarity value involves calculating two values: the local similarity value and the global similarity value.

1) Local similarity: The aim of similarity measurement is to quantify how much two objects resemble each other. Calculations for local similarity are conducted to obtain similarity values by comparing the attributes of a problem with those of a case. The local similarity is determined based on the characteristics of the data and its features [28].

- Numeric data type

$$f(s,t) = 1 - \frac{|s-t|}{R} \tag{1}$$

Here, *s* and *t* represent the values of the features under comparison, and *R* denotes the range of values associated with these features.

- Boolean data type

$$f(s,t) = \begin{cases} 1, if\ s = t \\ 0, if\ s \neq t \end{cases} \tag{2}$$

where s, t {true, false}

2) Global similarity: Global similarity is utilized for assessing the similarity between problems and cases on a case base. This study will compare the accuracy of systems using Minkowski Distance Similarity and Weighted Cosine Similarity.

- Minkowski distance simmilarity [29]

$$E(C_i, C_j) = \left[\frac{\sum_{k=1}^{n} w_k^r * |d_k(C_{ik}, C_{jk})|^r}{\sum_{k=1}^{n} w_k^r}\right]^{1/r} * T(C_j) * \frac{n(C_i, C_j)}{n(C_i)} \tag{3}$$

Here, $E(C_i, C_j)$ is the global similarity between target case (C_i) and source case (C_j), meanwhile w_k is the weight value of attribute k; $d_k(C_{ik}, C_{jk})$ is the local similarity value between target case attribute to k and source case attribute to k, and r is a Minkowski factor (positive integer); $T(C_j)$ is The confidence level of the case in the case base, $n(C_i, C_j)$ is the total attributes of the target case (C_i) that appear in the source case (C_j), and $n(C_i)$ is the total number of attributes in the target case (C_i)

- Weighted Cosine Similarity

$$\text{Weighted Cosine Similarity} = \frac{\sum_{i=1}^{n} w_i x_i y_i}{\sqrt{\sum_{i=1}^{n} w_i {x_i}^2}\ \sqrt{\sum_{i=1}^{n} w_i {y_i}^2}} \tag{4}$$

Here, w_i is the weight value of attribute i, x_i is the value of local similarity for the first object (*target case*), and y_i is the local similarity value of the second object (*source case*)

C. Implementation

The system will be developed as web-based software using Hypertext Markup Language (HTML)/Cascading Style Sheets (CSS) and Hypertext Preprocessor (PHP), with Apache serving as the webserver and MySQL handling the database.

D. Test Design

During system testing, the confusion matrix method is utilized to produce accuracy, recall (sensitivity), and precision values. The confusion matrix [30] acts as a concise result table, presenting the counts of true and false test data. This matrix facilitates a comparison between the actual values and the predicted results, allowing for the calculation of accuracy, prediction, and recall values, as depicted in Table I.

TABLE I. SYSTEM TESTING MATRIC

Predicted Values	Actual Values		
		Positive	Negative
	Positive	TP-True Positive	FP-False Negative
	Negative	FN-False Negative	TN-True Negative

The formula for calculating accuracy, precision, and recall [30] can be seen in equations five to seven.

- Accuracy: The extent of accuracy exhibited by the model in accurately performing the classification.

$$accuracy = \frac{TP+TN}{Total} \tag{5}$$

Precision: The degree of accuracy between the requested data and the predicted results from the model.

$$precision = \frac{TP}{TP+FP} \tag{6}$$

- Recall: The system's effectiveness in retrieving information.

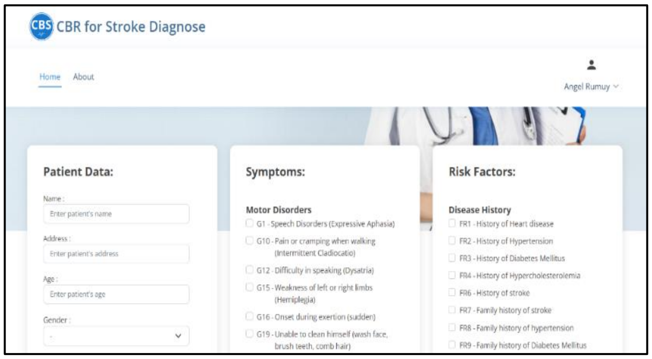
$$recall = \frac{TP}{TP+FN} \quad (7)$$

The test uses data from a research [31] in the form of medical record data from stroke patients during 2015-2016 at Dr Soetarto DKT Hospital, Yogyakarta consisting of 180 cases, where 30% of the cases, namely 54 cases, will be used as test data. The system underwent testing with various threshold values, specifically from 0.6 to 0.95. To find the highest accuracy value, the system will be tested using two different distance calculation methods, such as Minkowski Distance Similarity and Weighted Cosine Similarity methods. In a system that implements Minkowski Distance Similarity, a test is carried out on the Minkowski rank (r) to get the most optimal r value using different r values, started with r = 1 and continuing to increase by 1 until the resulting accuracy value does not show a significant difference.

III. Results

A. Stroke Diagnosis System

Research produces a system that can be used to diagnose stroke. The system has several interfaces, including an interface for carrying out diagnosis (see Fig. 3), an interface for displaying diagnosis results (see Fig. 4), and an interface for the system revision process (see Fig. 5).



CBR for Stroke Diagnose

Home About

Angel Rumuy

Patient Data:

Name : Enter patient's name

Address : Enter patient's address

Age : Enter patient's age

Gender : -

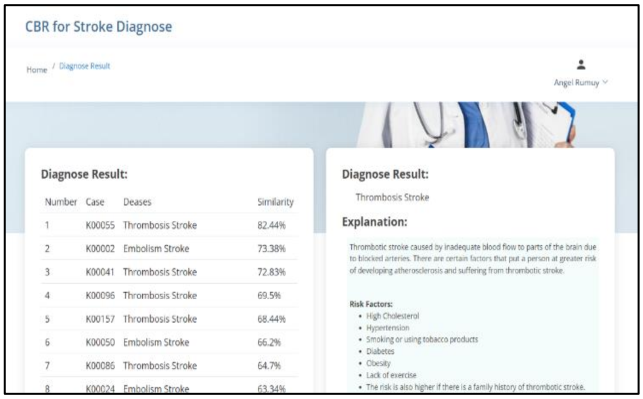
Symptoms:

Motor Disorders
☐ G1 - Speech Disorders (Expressive Aphasia)
☐ G10 - Pain or cramping when walking (Intermittent Cladiocatio)
☐ G12 - Difficulty in speaking (Dysatria)
☐ G15 - Weakness of left or right limbs (Hemiplegia)
☐ G16 - Onset during exertion (sudden)
☐ G19 - Unable to clean himself (wash face, brush teeth, comb hair)

Risk Factors:

Disease History
☐ FR1 - History of Heart disease
☐ FR2 - History of Hypertension
☐ FR3 - History of Diabetes Mellitus
☐ FR4 - History of Hypercholesterolemia
☐ FR6 - History of stroke
☐ FR7 - Family history of stroke
☐ FR8 - Family history of hypertension
☐ FR9 - Family history of Diabetes Mellitus

Fig. 3. Diagnosis page [27].



CBR for Stroke Diagnose

Home / Diagnose Result

Angel Rumuy

Diagnose Result:

Number	Case	Deases	Similarity
1	K00055	Thrombosis Stroke	82.44%
2	K00002	Embolism Stroke	73.38%
3	K00041	Thrombosis Stroke	72.83%
4	K00096	Thrombosis Stroke	69.5%
5	K00157	Thrombosis Stroke	68.44%
6	K00050	Embolism Stroke	66.2%
7	K00086	Thrombosis Stroke	64.7%
8	K00024	Embolism Stroke	63.34%

Diagnose Result:

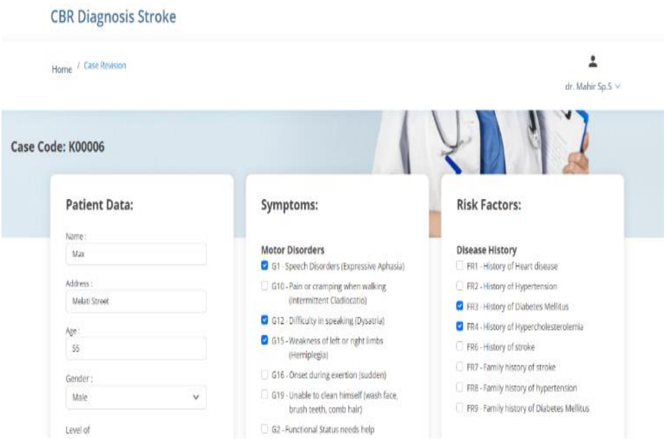
Thrombosis Stroke

Explanation:

Thrombotic stroke caused by inadequate blood flow to parts of the brain due to blocked arteries. There are certain factors that put a person at greater risk of developing atherosclerosis and suffering from thrombotic stroke.

Risk Factors:
- High Cholesterol
- Hypertension
- Smoking or using tobacco products
- Diabetes
- Obesity
- Lack of exercise
- The risk is also higher if there is a family history of thrombotic stroke.

Fig. 4. Diagnosis result page [27].



CBR Diagnosis Stroke

Home / Case Revision

dr. Mahir Sp.S

Case Code: K00006

Patient Data:

Name : Max

Address : Melati Street

Age : 55

Gender : Male

Level of

Symptoms:

Motor Disorders
☑ G1 - Speech Disorders (Expressive Aphasia)
☐ G10 - Pain or cramping when walking (Intermittent Cladiocatio)
☑ G12 - Difficulty in speaking (Dysatria)
☑ G15 - Weakness of left or right limbs (Hemiplegia)
☐ G16 - Onset during exertion (sudden)
☐ G19 - Unable to clean himself (wash face, brush teeth, comb hair)
☐ G2 - Functional Status needs help

Risk Factors:

Disease History
☐ FR1 - History of Heart disease
☐ FR2 - History of Hypertension
☑ FR3 - History of Diabetes Mellitus
☑ FR4 - History of Hypercholesterolemia
☐ FR6 - History of stroke
☐ FR7 - Family history of stroke
☐ FR8 - Family history of hypertension
☐ FR9 - Family history of Diabetes Mellitus

Fig. 5. Expert revise page [27].

On the diagnosis page, the user enters the symptoms experienced by a patient. This page is the input page for the system. Based on input from the user, the system calculates the similarity of the user input with the case dataset that the system already has. Instances of case representation are illustrated in Table II. When the system identifies similar cases (with similarity exceeding the threshold value), it will present a diagnosis results page. In the absence of cases resembling the symptoms entered by the user, the system will store the case data, and experts can review new cases through the system's revision page.

TABLE II. Example of Case Representation [26]

Base Case		
Patient Code:		K00007
General Condition:		
1	Age	60
2	Gender	Male
3	Awareness	Compo Mentis
Symptom:		
G1	Confusion	No
G3	Trouble balancing	No
Gn	n-th symptom	…
Risk Factor:		
FR1	History of heart disease	No
FR2	History of hypertension	Yes
FRn	n-th risk factor	…
Diagnosis:	Embolism Stroke	

C. System Testing Results

System testing involves the calculation of accuracy, precision, and recall using Eq. (5), (6), and (7). The testing process utilizes 30% of the case data as test data, consisting of 54 cases. To enhance efficiency, an automation script is employed. This script logs into the system, automatically inputs patient data, symptoms, and risk factors based on the test data, and records the system results.

The test outcomes for the system utilizing the Minkowski Distance Similarity method, along with the Confusion Matrix, are detailed in Table III. This table depicts different levels of accuracy, sensitivity, and recall corresponding to each threshold value and *r* value. The peak accuracy is achieved with a threshold value of 75 and *r* values of three and four, resulting in an accuracy rate of 88.89%.

Table IV displays the outcomes of system testing utilizing the Weighted Cosine Similarity method and the confusion matrix. The highest accuracy is achieved with threshold values of 75 and 80, yielding an accuracy percentage of 83.33%. The accuracy value signifies the system's ability to diagnose correctly, with higher accuracy indicating more precise diagnosis results or solutions provided by the system.

TABLE III. MINKOWSKI DISTANCE SIMILARITY SYSTEM TEST RESULTS

Threshold	Nilai R	Accuracy (%)	Recall (%)	Precision (%)
5	1	72,22	78	90,7
10	1	72,22	78	90,7
15	1	72,22	78	90,7
20	1	72,22	78	90,7
25	1	72,22	78	90,7
30	1	72,22	78	90,7
35	1	72,22	78	90,7
40	1	74,07	78	92,86
45	1	74,07	74	97,37
50	1	81,48	80	100
55	1	85,19	84	100
60	1	85,19	84	100
65	2	85,19	84	100
70	2	87,04	86	100
75	3 & 4	88,89	88	100
80	8	77,78	76	100
85	12	62,96	60	100
90	12	38,89	34	100
95	12	37,04	32	100

TABLE IV. WEIGHTED COSINE SIMILARITY SYSTEM TEST RESULTS

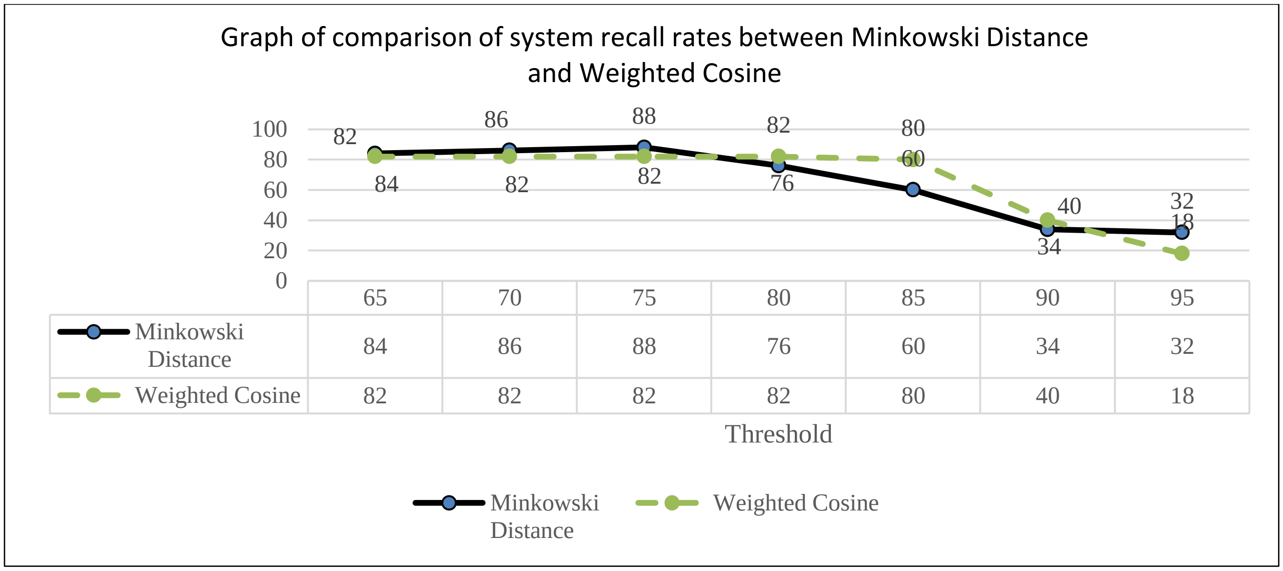
Threshold	Accuracy (%)	Recall (%)	Precision (%)
5	75,93	82	91,11
10	75,93	82	91,11
15	75,93	82	91,11
20	75,93	82	91,11
25	75,93	82	91,11
30	75,93	82	91,11
35	75,93	82	91,11
40	75,93	82	91,11
45	75,93	82	91,11
50	77,78	82	93,18
55	77,78	82	93,18
60	77,78	82	93,18
65	79,63	82	95,35
70	81,48	82	97,62
75	83,33	82	100
80	83,33	82	100
85	81,48	80	100
90	44,44	40	100
95	24,07	18	100

IV. DISCUSSION

Accuracy calculates all actual predicted values without specificity for each label, so a higher accuracy doesn't necessarily indicate good performance in predicting specific labels. Therefore, recall and precision values are crucial. Recall assesses the system's success in retrieving information, with higher values indicating better identification of positive cases.

Fig. 6 depicts the system's recall rate at a threshold of more than or equal 75 using the Minkowski Distance method, which surpasses the recall rates of the system using the Jaccard Coefficient method without indexing and the Weighted Cosine method. The Minkowski Distance method achieves the highest recall value at 88%. Precision, a metric measuring the accuracy of positive predictions, is highest in the Minkowski Distance method when applying a threshold value of more than or equal 50, reaching 100% (see Fig. 7).

Across the three tested methods, the Minkowski Distance approach with a threshold value of 75 and *r* values of three or four consistently produces the highest levels of accuracy, recall, and precision. In a system designed to detect high-risk diseases such as stroke, recall is particularly crucial, as a low recall value implies misdiagnosing some patients with stroke as healthy, leading to serious risks. Therefore, the optimal configuration for the expert system for stroke diagnosis is the Minkowski Distance Similarity method with a threshold value of 75 at r = 3 or r = 4, achieving a system accuracy rate of 88.89%, a recall of 88%, and a precision of 100%.



	65	70	75	80	85	90	95
Minkowski Distance	84	86	88	76	60	34	32
Weighted Cosine	82	82	82	82	80	40	18

Fig. 6. Graph of comparison of system recall rates between Minkowski Distance, Weighted Cosine and Jaccard Coefficient.

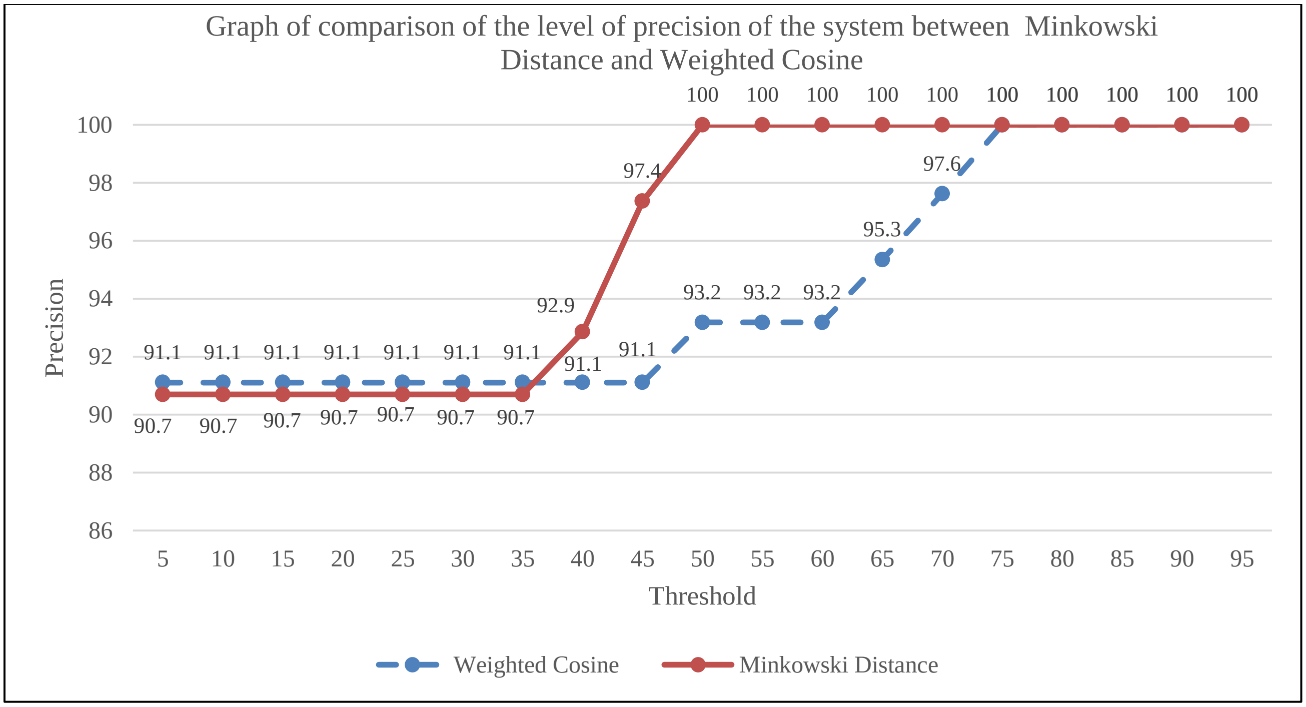


Fig. 7. Graph of comparison of the level of precision of the system between Minkowski Distance and Weighted Cosine.

A. Limitation

This research is a follow-up research that uses the same dataset from the work of Nelson and the team [26]. This study has not added a dataset with the latest cases for stroke diagnosis. Further research can be carried out by collecting and adding a dataset of cases in the last five years.

V. CONCLUSION

This research is a follow-up study that uses a dataset of stroke cases. Research focuses on the effectiveness of algorithms for stroke diagnosis. The system developed is an expert system with a CBR approach. The study focuses on finding the most effective algorithm for diagnosing stroke. In previous research, the Jaccard Coefficient algorithm was applied with an accuracy level of 81.67%, and Minkowski Distance Similarity was applied with an accuracy of 88.89%. In this research, the Weighted Cosine algorithm was applied, resulting in an accuracy of 83.33%. Through the comparison of the applications of the three algorithms, it becomes apparent that the Minkowski Distance Similarity algorithm exhibits a superior level of accuracy and sensitivity (or recall) in contrast to systems utilizing the Jaccard Coefficient method and the Weighted Cosine method. When the threshold is set at 75 or above, the system attains an accuracy rate of 88.89%, along with a recall of 88%. In comparison to the Weighted Cosine method alone, the precision level is 100%.

Further research is being carried out to develop a rule generator to automate the formation of a knowledge base in a rule-based system format. The anticipated outcome of this research is an enhancement in the efficiency of tracking case times for decision-making. Apart from this, updates to the stroke case dataset must also be carried out to enrich the case data

ACKNOWLEDGMENT

We express our gratitude to the Faculty of Information Technology and the Institute for Research and Community Service (LPPM) at Duta Wacana Christian University for providing the funding for both the research and its publication.

REFERENCES

[1] "Apa itu Stroke," http://p2ptm.kemkes.go.id/infographic-p2ptm/stroke/apa-itu-stroke. 2018.

[2] P. J. F. Lucas and L. C. Van Der Gaag, Principles of expert systems. Singapore: Addison-Wesley Publishing, 1991. [Online]. Available: https://www.researchgate.net/publication/220694050

[3] A. Abraham, "Rule-based Expert Systems," in Handbook of Measuring System Design, P. H. Sydenham and R. Thorn, Eds. John Wiley & Sons, Ltd., 2005, pp. 909–919.

[4] I. Gunawan and Y. Fernando, "Sistem Pakar Diagnosa Penyakit Kulit pada Kucing Menggunakan Metode Naive Bayes Berbasis Web," J. Inform. dan Rekayasa Perangkat Lunak, vol. 2, no. 2, pp. 239–247, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika/article/view/927/380

[5] F. Anjara and A. A. Jaharadak, "Expert System for Diseases Diagnosis in Living Things: A Narrative Review," J. Phys. Conf. Ser., vol. 1167, no. 1, 2019, doi: 10.1088/1742-6596/1167/1/012070.

[6] H. Henderi, F. Al Khudhorie, G. Maulani, S. Millah, and V. T. Devana, "A Proposed Model Expert System for Disease Diagnosis in Children to Make Decisions in First Aid," INTENSIF J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf., vol. 6, no. 2, pp. 139–149, 2022, doi: 10.29407/intensif.v6i2.16912.

[7] A. Andriani, A. Meyliana, Sardiarinto, W. E. Susanto, and Supriyanta, "Certainty Factors in Expert System to Diagnose Disease of Chili Plants," 2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018, no. Citsm, 2019, doi: 10.1109/CITSM.2018.8674264.

[8] B. Basiroh and S. W. Kareem, "Analysis of Expert System for Early Diagnosis of Disorders During Pregnancy Using the Forward Chaining Method," Int. J. Artif. Intell. Res., vol. 5, no. 1, pp. 44–52, 2021, doi: 10.29099/ijair.v5i1.203.

[9] B. A. Wijaya and J. P. Tanjung, "An Expert System For Diagnosis Eye Diseases On Human Using Certainty Factor Method Based Web," SinkrOn, vol. 5, no. 1, pp. 78–83, 2020, doi: 10.33395/sinkron.v5i1.10579.

[10] X. Huang et al., "A Generic Knowledge Based Medical Diagnosis Expert System," ACM Int. Conf. Proceeding Ser., vol. 1, no. 1, pp. 462–466, 2021, doi: 10.1145/3487664.3487728.

[11] I. Setiawan and M. Batara, "Expert System Design to Diagnose Pests and Diseases on Local Red Onion Palu Using Bayesian Method," BAREKENG J. Math. Its Appl., vol. 17, no. 1, pp. 371–382, 2023.

[12] M. M. Richter and R. O. Weber, Case-Based Reasoning. New York: Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-40167-1.

[13] I. Nugraha and M. Siddik, "Penerapan Metode Case Based Reasoning (CBR) Dalam Sistem Pakar Untuk Menentukan Diagnosa Penyakit Pada Tanaman Hidroponik," J. Mhs. Apl. Teknol. Komput. dan Inf., vol. 2, no. 2, pp. 91–96, 2020, [Online]. Available: https://www.ejournal.pelitaindonesia.ac.id/JMApTeKsi/index.php/JOM/article/view/575/387

[14] I. Y. Subbotin and M. G. Voskoglou, "Applications of fuzzy logic to Case-Based Reasoning," vol. 11, pp. 7–18, 2012, [Online]. Available: https://www.researchgate.net/publication/223129950

[15] A. S. Soroto, A. Fuad, S. Lutfi, J. J. Metro, and K. T. Selatan, "Penerapan Metode Case Based Reasoning (CBR) untuk Sistem Penentuan Status Gunung Gamalama," J. Inform. dan Komput., vol. 02, no. 2, pp. 70–75, 2018.

[16] A. Yuli Vandika and A. Cucus, "Sistem Deteksi Awal Penyakit TBC dengan Metode CBR," Pros. Semin. Nas. Darmajaya, 2017.

[17] J. Dou et al., "Automatic Case-Based Reasoning Approach for Landslide Detection: Integration of Object-Oriented Image Analysis and a Genetic Algorithm," Remote Sens., vol. 7, no. 4, pp. 4318–4342, 2015, doi: 10.3390/rs70404318.

[18] F. Tempola, A. Musdholifah, and S. Hartati, "Case Based Reasoning Untuk Penentuan Kelayakan Mahasiswa Penerima Beasiswa," 2015.

[19] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," J. Inform. J. Pengemb. IT, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.

[20] J. Firdaus, I. Y. Purbasari, and H. P. Swari, "Implementasi Case-Based Reasoning pada Sistem Prediksi Masa Studi dan Predikat Kelulusan Mahasiswa (Studi Kasus : Fakultas Ilmu Komputer, UPN 'Veteran' Jawa Timur)," 2020.

[21] R. Adawiyah, "Implementasi Metode Minkowsky Distance untuk Deteksi Kelahiran Bayi Prematur Berbasis Case-Based Reasoning," J. Inform. dan Komputer) Akreditasi KEMENRISTEKDIKTI, vol. 3, no. 1, 2020, doi: 10.33387/jiko.

[22] A. Mubarak, M. Salmin, A. Fuad, and S. Do Abdullah, "Penalaran Berbasis Kasus Untuk Diagnosis Penyakit Malaria Dengan Menggunakan Metode Minkowsky Distance," J. Ilm. Ilk. - Ilmu Komput. Inform., vol. 5, no. 1, 2022, doi: 10.47324/ilkominfo.v4i3.136.

[23] M. E. Soinbala, D. Rony Sina, and M. Boru, "Case Based Reasoning untuk mendiagnosis Gizi Buruk pada Anak Usia 0-5 Tahun Menggunakan Metode Cosine Similarity," J-ICON, vol. 7, no. 1, pp. 67–71, 2019.

[24] M. Zainuddin, K. Hidjah, and W. Tunjung, "Penarapan Case-Based Reasoining (CBR) untuk Mendiagnosis Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor," CITISEE, 2016.

[25] I. Warman, "Sistem Pakar Identifikasi Penyakit Tanaman Padi Menggunakan Case-Based Reasoning," 2017.

[26] R. Nelson, A. Harjoko, and A. Musdholifah, "Case-Based Reasoning for Stroke Disease Diagnosis," IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 12, no. 1, p. 33, 2018, doi: 10.22146/ijccs.26331.

[27] A. Rumuy, R. Delima, K. P. Saputra, and J. Purwadi, "Application of the Minkowski Distance Similarity Method in Case-Based Reasoning for Stroke Diagnosis," JUITA J. Inform., vol. 11, no. 2, pp. 323–332, 2023, [Online]. Available: https://jurnalnasional.ump.ac.id/index.php/JUITA/article/view/18582/pdf

[28] M. K. Jha, D. Pakhira, and B. Chakraborty, "Diabetes Detection and Care Applying CBR Techniques ," IJSCE, vol. 2, no. 6, 2013.

[29] E. Faizal and H. Hamdani, "Weighted Minkowski Similarity Method with CBR for Diagnosing Cardiovascular Disease," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 12, 2018, doi: 10.14569/IJACSA.2018.091244.

[30] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.

[31] R. Nelson, A. Harjoko, and A. Musdholifah, "Rekam Medik Stroke.xlsx." 2018.

Optimal Cluster Head Selection in Wireless Sensor Network via Combined Osprey-Chimp Optimization Algorithm: CIOO

Vikhyath K B[1], Achyutha Prasad N[2]

Research Scholar, Department of Computer Science and Engineering, East West Institute of Technology, Bengaluru
Visvesvaraya Technological University, Belagavi, India - 590018[1]
Research Supervisor-Department of Computer Science and Engineering, East West Institute of Technology, Bengaluru
Visvesvaraya Technological University, Belagavi, India – 590018[2]

***Abstract*—The development of Wireless Sensor Network (WSN) has gained significant attention for smart systems due to their potential use in a wide range of areas. WSN consists of tiny, independently arranged sensor nodes that run on batteries. The resources and energy usage for sensor nodes are the major factors. Particularly, the unbalanced nodes' raises the energy use and reduces the network life-span. Energy efficiency in WSN cluster head selection remains a challenging task. The best method has been developed for reducing node energy consumption is clustering. However, the current clustering strategy failed to properly allocate the energy needs of the nodes without considering energy features, node quantity, as well as adaptability. Hence, there is need for advanced clustering process with new optimization tactics, and accordingly, a new cluster-head selection model in WSN is proposed in this work. Initially, the clustering process is done by the k-means algorithm. The Cluster Head (CH) selection is the subsequent progress under the consideration of node's energy, distances, delays, and risks as well. A novel CIOO (Chimp Integrated Osprey Optimization) algorithm combining the Osprey and Chimp optimization algorithm is proposed for Cluster Head Selection (CHS). Finally, the performance of proposed model is evaluated over the conventional methods.**

Keywords—Wireless sensor network; clustering; cluster head; cluster head selection; chimp optimization; osprey optimization

I. Introduction

In recent years, a wide range of application domains, including military operations, medical care, smart cities, renewable electric power and energy, with monitoring the environment, have benefited significantly from the technological innovation and research carried out on WSNs. WSN is made up of BS and many sensor nodes [1, 2]. These are tiny independent devices with many limitations, including battery life, low processing power, and short communication distance [3]. To detect or sense variations in environmental factors like temperatures, movement, stress, moisture, vibration, noise, etc., these nodes are scattered geographically over a vast region. Nodes are effective enough to speed up data transfer via wireless networks. One of the fundamental technologies of substantial WSN, data collecting has attracted major academic interest [4, 5]. The data is sent to a data sinks or BS after collection. Sensor nodes are periodically placed in dangerous locations; and in these situations, replacing the batteries is not feasible. The sensor nodes use more energy since they constantly transmit and receive data. When nodes consume too much energy, they fail and can no longer be repaired or recharged with another battery source [6]. Therefore, balancing the nodes energy is the main issue in WSN [7, 8]. In order to increase the longevity and efficiency of the network, it is essential to allocate node energy correctly and optimize node energy consumption. Networking solutions related to sustainability and efficacy are being explored for energy concerns through hierarchical clustering algorithms [9, 10]. A familiar cluster-based routing methods like Low Energy Adaptive Clustering Hierarchy (LEACH) [11], Centralized-LEACH, and Improved-LEACH significantly aids in optimizing node energy towards the network longevity [12, 13]. However, such protocols are constrained as they choose an arbitrary applicant for the CH without taking energy variables into account, and not succeed to accomplish optimal formation of clusters and CH. This resulting in exhibiting greater communication costs, route by a single hop, and consume a greater amount of energy [14, 15] leading to network lifetime extension. Also, more optimization algorithms do involve in this CHS since from its evaluation, however, few existing optimization techniques struggle to preserve the level of investigation and extraction of choosing CH since they concentrate on particular search domains. Hence, proposing advanced optimization algorithms for optimal clustering is needed, and accordingly, this paper intends to prepare a new cluster-based routing model. The main contribution of the proposed model is given below:

- Proposing a new CIOO algorithm that combines both osprey optimization algorithm and chimp optimization algorithm for the CH selection, and also determining an improved trust evaluation as the constraint for routing.
- The proposed algorithm is subjected for evaluation with the conventional algorithms to emphasis its effectiveness.

The remaining part of this research paper is structured as below. Literature review is shown in Section II. The proposed cluster head selection method is illustrated in Section III. The results and discussion are highlighted in Section IV. Conclusion and future scope of the recommended model is provided in Section V.

II. Literature Review

The latest advancements in cluster head selection in WSN are examined in this section. The authors in [16] stated that one of the most important techniques used to extended the lifespan of WSNs operated by batteries was clustering routing. However, the majority of currently used clustering techniques fail to make use of the node redundancies in WSNs, resulting in significant energy waste. The authors in [17] introduced a technique for cluster head selection in multilevel topological control that makes use of fuzzy grouping pre-processing and optimization of particle swarms. The authors in [18] created a distributed CH selection approach whereby distances among sensors and a base station were taken into account to ensure continuous optimization of energy usage among the sensors. Authors in [19] introduced a brand-new technique termed clustering-based Cluster-Head Selection Scheme with Power Control (CHESS-PC) in PSN. Authors in [20] explored the Dynamic Cluster Head Selection Methods (DCHSM) for WSN to address the problems of unreasonable cluster leader selection, which in turn results in imbalanced consumption of energy and inconsistent coverage in the cluster communications.

Authors in [21] presented an approach called LEACH to control the unpredictability that occurs in clustering algorithms. With this strategy, the cluster head count was stabilized. Authors in [22] proposed a Hausdorff clustering technique, a new static kind of cluster approach based on the arrangement of sensor nodes that also improves network interaction and effectiveness. Authors in [23] proposed the Cluster Head Selection by Randomness with Data Recovery (CHSRDR) approach, which was a new way for WSN to select the cluster head with recovered data and maintained inside the cluster. Authors in [24] created a Firefly algorithm for choosing the cluster head in a manner which was sufficiently near to the base station as well as the sensor nodes. As a result, the period of delay was significantly compact, increasing the information packets' transmission speed. Authors in [25] introduced a Particle Swarm Optimization (PSO) technique for creating energy-aware clusters by selecting the best cluster heads. The PSO has ultimately reduced the expenditure of determining the perfect place for the CH nodes.

Authors in [26] created a Fuzzy Based Balanced Cost CH Selection Algorithm (FBECS) that takes into consideration the remaining energy, distance toward the sink, and population of the node in the area as inputs to the fuzzy decision-making method. Authors in [27] introduced Hyper Exponential Reliability Factor Based Cluster Head Election (HRFCHE), an integrated prediction technique for energy and trust evaluation for extending the network's lifespan. By extending the network lifetime and decreasing consumption of energy by 28% and 34%, respectively, compared to cluster head selection systems used as a benchmark, the outcomes of HRFCHE have implied greater efficiency. Authors in [28] concentrated on the effective operation of WSN applications, proving that the energy-efficient operation of the sensors became an important framework for extending the lifespan of the network. Authors in [29] implemented a Fuzzy-TOPSIS method in Select CH effectively and efficiently in order to optimize the WSN lifetime.

Even though the literature survey deals with efficient cluster head selection methods, there is still scope to improve the overall lifetime of the network and increase the number of alive nodes. These gaps can be addressed by developing new multi-constraint-based optimal cluster head selection methods.

III. Proposed Chimp Integrated Osprey Optimization Algorithm for Optimal Cluster Head Selection in WSN

The two main obstacles in wireless sensor networks are selecting the right cluster head and energy-related constraints. In order to overcome these obstacles and extend the network's lifespan, optimized algorithms are crucial. Clustering is the most critical process for increasing network longevity in WSNs. Sensor nodes are organized into clusters, and each cluster is assigned a CH. The CHs take data from the nodes in each cluster and forward it to the base station. Choosing the suitable CH in WSNs is a major challenge. Four criteria's energy, delay, distance, and security are used to establish a novel cluster head selection framework in our proposed model. The suggested model is summarized as follows:

- Initially, the clustering procedure groups the sensor nodes together. The k-means clustering algorithm is used in our suggested work to perform the clustering procedure.
- After the cluster's development, the cluster head is selected by CIOO algorithm which combines both osprey optimization algorithm and chimp optimization algorithm.
- The new CIOO algorithm is executed with four parameters like energy, distance, delay, risk.

A. *Clustering via k-means Clustering*

The research community has focused heavily on clustered to address the energy, scalability and lifetime problems of WSNs. Clustering algorithms restrict connection to a local area and utilized forwarding nodes to send the essential data to the reaming nodes of the network. Generally, the cluster members do interact with the CH, and the CH collects and combines the information gathered to save the energy. Here, the cluster heads can additionally create an additional layer of clusters between them. The k-means procedure [30], which offers straightforward, extremely dependable, fast-convergent repetitions & re-clustering throughout failure conditions, is a well-liked centralized as well as spread probabilistic partitional clustering method. The clustering process in the k-means algorithm is largely dependent on Euclidian distances. The procedure for k means algorithm is explained in step-by-step procedure given below.

- Group the nodes into 'k' clusters, take 'k' centroids and arrange them initially at random locations.
- Calculate the nearest centroid by calculating the Euclidian distance among every node as well as the entire center. Initial clusters are created by this process, 'k'.
- Recalculating the locations of centroids in every cluster and any changes to be checked from the prior calculation.

- If any changes in the centroid position, then proceed to Step 2; else the clustering process is complete and finalized the clusters.

This involves grouping the nodes into 'k' clusters, and selecting the CHs for every cluster is by the hybrid optimization technique like CIOO technique, which is explained in the subsequent section.

B. Optimal Cluster Head Selection via CIOO Algorithm

After the clustering process, optimal cluster head is selected depends on the consideration of nodes energy, distance like inter-cluster and intra cluster distances, delay and Risk factors by using CIOO algorithm. The calculation of distance, delay and risk of the clustering nodes are explained as follows. Fig. 1 shows the architecture of CH selection via CIOO algorithm.

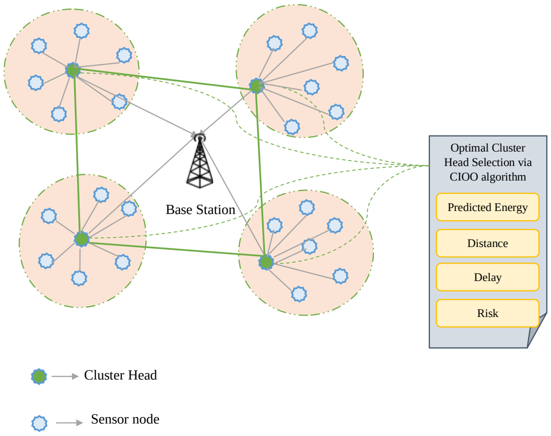


Fig. 1. CH selection framework by CIOO algorithm.

1) Distance (Δ_{CH})*:* The spacing among nodes in identical and separate clusters is measured by distance (Δ_{CH}). Here, Inter-cluster distance and intra-cluster distance are the two forms of distance that are computed. The Euclidean, Manhattan, and Chebychev distance formulas are utilized for the calculation of the distance among the cluster's nodes. Let R, S be the clusters, and their distance of each node in the cluster be $|R|$ and$|S|$, respectively. There are two distances determined by inter-cluster and intra-cluster distances are Average Linkage Distance and Complete Diameter Distance, respectively.

a) Average Linkage Distance (Inter cluster distance): Eq. (1) defines the calculation of averaged linkage distance that is the average distance between all of the nodes in two distinct clusters, where X represents the node in cluster, R and Y indicates the nodes in cluster S.

$$\Delta(R,S) = \frac{1}{|R||S|} \sum_{\substack{x \in R \\ y \in S}} d(\mathrm{x},\mathrm{y}) \tag{1}$$

b) Complete Diameter Distance (Intra cluster distance): Using Eq. (2), the whole diameter separation is determined as the spacing between two nodes in the same cluster (CH and Sensor node) which are located the furthest apart from each other.

$$\Delta(R) = \max\{d(x,y)\} \tag{2}$$

2) Delay (D_{CH})*:* When there do not exist accessible node for delaying the data, a delay constraint in a WSN is defined as the time interval between the dispersed data of one-time unit and another. As a result, the delay is more closely related to the delay of the probabilistic transmitting system. The mathematical computation of delay, which is the ratio of distance and speed, is shown in Eq. (3), where D' denotes the distance and S' denotes the speed.

$$Delay\ (D_{CH}) = \frac{D'}{S'} \tag{3}$$

3) Risk (f_{risk})*:* The various elements of security methods such as risky mode, the γ-risky mode and the security mode [31] that are explained below.

a) Risky mode: This strategy selects a present CH and accepts all risks to promote an ideal CHS. As a result, choosing CH is regarded as choosing an aggressive mode.

b) γ *-risky mode:* The setting of the Cluster Head which can handle the maximum of γ-risk level is selected based on the "γ-risky mode." Therefore, a likelihood metrics having a value between 0 and 1 of 100% that shows the secure and risky modes is represented by the symbol γ.

c) Security mode: This option supports the CH, which fulfills with security standards in security mode. The variables R^{sd} and R^{sr} indicates the security demand and the security rank pertaining to CH selection. The node is considered to be CH if $R^{sd} \leq R^{sr}$. Security constraints is determined as shown in (4). Additionally, "the risk should be 50% below if the chosen CH achieves the stated $R^{sd} < R^{sr}$. If the condition $0 < R^{sd} - R^{sr} \leq 1$ is true, the selecting process would proceed as planned rather than being delay in $1 < R^{sd} - R^{sr} \leq 2$. The state $2 < R^{sd} - R^{sr} \leq 5$ continues carrying out the related function as the CHS process unable to be completed.

$$f_{risk} = \begin{cases} 0 & \text{, if } R^{sd} - R^{sr} \leq 0 \\ 1 - e^{\frac{\left(R^{sd} - R^{r}\right)}{2}} & \text{if } 0 < R^{sd} - R^{sr} \leq 1 \\ 1 - e^{\frac{3\left(R^{dd} - R^{r}\right)}{2}} & \text{, if } 1 < R^{sd} - R^{sr} \leq 2 \\ 1 & \text{,if } 2 < R^{sd} - R^{sr} \leq 5 \end{cases} \tag{4}$$

C. Objective Function for CH Selection

The objective function for CH selection is calculated depends on the minimum fitness function of constraints like nodes energy, Delay, Distance and Risk. In the proposed work, two objective functions are considered, the first objective function is based on the consideration of Risk, Distance, Delay and the second objective function is based on the consideration of Energy. The fitness and the first objective function $f_1'$of Risk, Delay and Distance for CH selection is evaluated in Eq. (5) and Eq. (6), which obtained the minimum value of Risk, Distance and Delay, where,w_1, $w_2$and w_3 indicates the weight of Risk, Distance and Delay respectively. The summation of the total weight is represented as $\sum w_i = 1$. Then the second objective function f_2' of cluster head selection under the consideration of nodes energy is determined as per Eq. (7)

and Eq. (8). Here, w_4 denotes the weight of energy, which obtain maximum in energy consumption.

$$\text{Fit}= \text{Minimum}\left(w_1 * f_{\text{risk}} + w_2 * \Delta_{CH} + w_3 * D_{CH}\right) \quad (5)$$

$$\text{obj}(f_1') = \text{Minimum}(\text{ Fit }) \quad (6)$$

$$\text{Fit } = \text{Minimum}(w_4 * (1 - E)) \quad (7)$$

$$\text{obj}(f_2') = \text{Minimum}(\text{ Fit }) \quad (8)$$

Finally, combined the objective function is mathematically defined in Eq. (9), where, m denotes the parameter in the range of [0, 1], f_1' denotes the overall objective function for the CH selection, $f_1'$and f_2' are the two objective functions.

$$F' = m * f_1' + (1 - m) * f_2'; \;\; 0 < m < 1 \quad (9)$$

D. Solution Encoding for CH Selection

The Solution given to the CIOO algorithms are nodes. The lower bound value is fixed as 1 and the upper bound value is n, where, n denotes the number of sensor nodes. Here, the population size is allocated as 10. The flowchart in Fig. 2 shows the CIOO algorithm implementation processes.

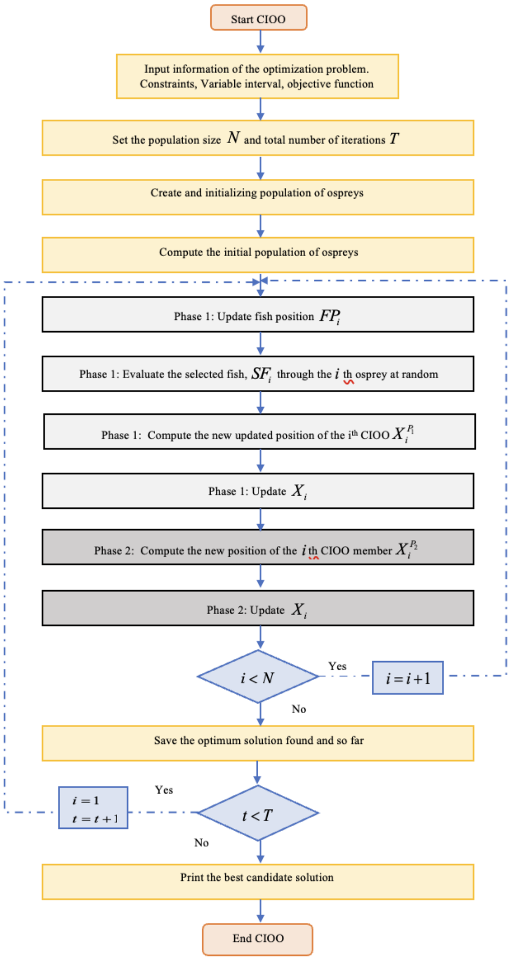

Start CIOO
Input information of the optimization problem. Constraints, Variable interval, objective function
Set the population size N and total number of iterations T
Create and initializing population of ospreys
Compute the initial population of ospreys
Phase 1: Update fish position FP_i
Phase 1: Evaluate the selected fish, SF_i through the i th osprey at random
Phase 1: Compute the new updated position of the i^th CIOO $X_i^{P_1}$
Phase 1: Update X_i
Phase 2: Compute the new position of the i th CIOO member $X_i^{P_2}$
Phase 2: Update X_i
$i < N$
Yes
$i = i + 1$
No
Save the optimum solution found and so far
$i = 1$
$t = t + 1$
Yes
$t < T$
No
Print the best candidate solution
End CIOO

Fig. 2. Flowchart of CIOO algorithm.

IV. Results and Discussion

A. Simulation Procedure

The simulation of the proposed cluster-based routing in Wireless Sensor Networks (WSN) was conducted using MATLAB, with the MATLAB version being "Matlab R2018a."Further, the processor utilized was "Intel(IR) Core(TM) i5-1035G1 CPU @1.00GHz 1.19 GHZ" and the system had a total installed RAM size of "20.0GB," with "19.7GB" of it being usable.

B. Performance Analysis

Additionally, the performance of both the CIOO and conventional approaches was evaluated across various metrics, including Distance, Total Packets Transmitted to the Base Station (BS), Residual Energy, Delay, Alive Nodes, and Risk. Furthermore, the CIOO method was compared with state-of-the-art approaches such as DMOSC-MHRS [32] and PSO [33]. Additionally, a comparative analysis was conducted between the CIOO method and traditional algorithms, including GOA [34], SMO [35], BOA [36], COA [37], and OOA [38]. The network setup and energy model were illustrated in Fig. 3.

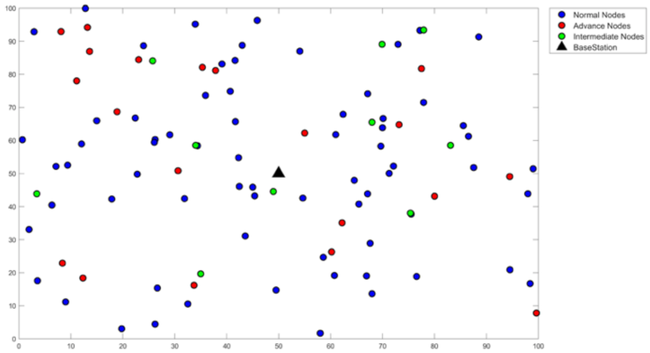


Fig. 3. Network setup.

C. Analysis on Delay and Distance

Fig. 4 and Fig. 5 provide an explanation of the delay and distance evaluation in comparison to PSO [33], DMOSC-MHRS [32], GOA [34], SMO [35], BOA [36], COA [37], and OOA [38] for the optimal selection of cluster heads. Additionally, this analysis is conducted while varying the number of rounds (500, 1000, 1500, and 2000). The objective for achieving optimal cluster head selection is to minimize both delay and distance ratings. Interestingly, at the 1000th round, the CIOO method exhibited the highest delay and distance values. However, as the number of rounds increased beyond 1000, there was a noticeable decrease in both delay and distance rates. Mainly, at the round 2000, the CIOO method achieved an impressively low delay value of 2134s. In contrast, traditional schemes recorded notably higher delay values, such as, PSO [33] =2856s, DMOSC-MHRS [32] =2150s, GOA [34] =2598s, SMO [35] =2342s, BOA [36] =2831s, COA [37] =2797s, and OOA [38] =2782s, respectively. In addition, the distance rate attained by the CIOO scheme is 8.732×104 at the round 1500, whereas the PSO [33], DMOSC-MHRS [32], GOA [34], SMO [35], BOA [36], COA [37] and OOA [38] resulted in greater distance ratings. As a result, the CIOO method employs a hybrid optimization strategy that combines

OOA [38] and COA [37] to achieve optimal cluster head selection in WSN. This method consistently achieves dependable results by efficiently decreasing both delay and distance metrics.

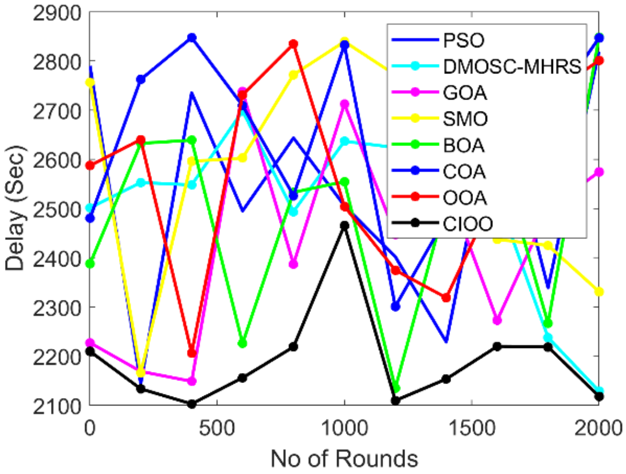


Fig. 4. Delay validation of CIOO with conventional algorithms.

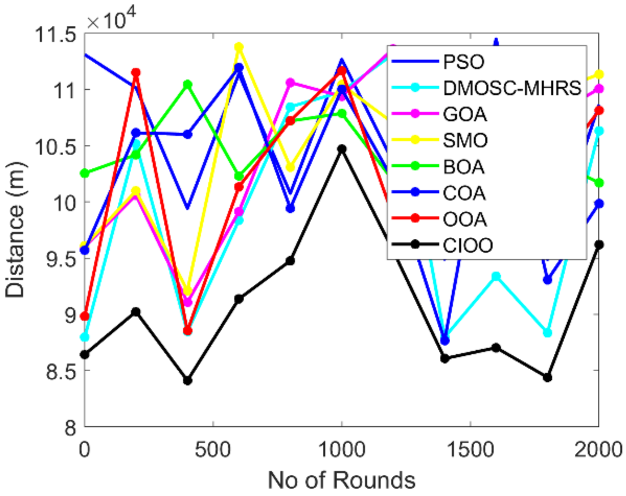


Fig. 5. Distance validation of CIOO with conventional algorithms.

D. Analysis on Alive Nodes

Fig. 6 presents a comparative analysis of the number of alive nodes in the CIOO approach versus PSO [33], DMOSC-MHRS [32], GOA [34], SMO [35], BOA [36], COA [37], and OOA [38] for cluster-based routing in WSN. In the pursuit of achieving optimal cluster-based routing in WSN, the primary goal is to maximize the number of nodes that remain active or "alive." During the initial round, both the CIOO and conventional approaches achieved the highest number of alive nodes. Nevertheless, as subsequent rounds progressed, the number of surviving nodes declined. Nevertheless, it's worth noting that the CIOO method consistently outperformed the conventional approaches by maintaining a higher number of active nodes. Significantly, the CIOO method achieved the highest number of active nodes, reaching 42 at round 2000. This count is notably superior to the numbers achieved by PSO [33], DMOSC-MHRS [32], GOA [34], SMO [35], BOA [36], COA [37], and OOA [38].

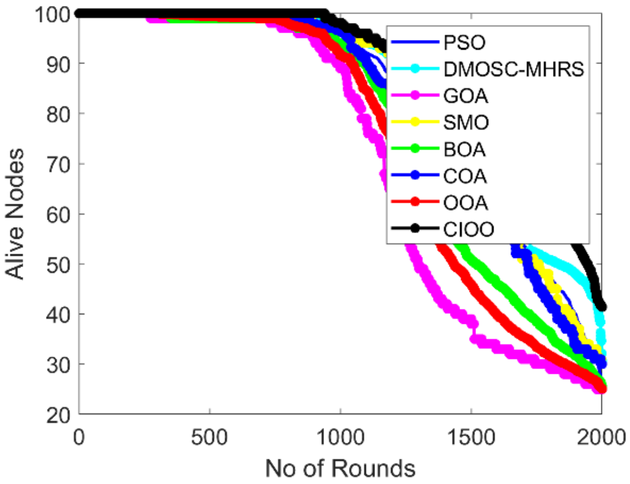


Fig. 6. Validation of CIOO with conventional algorithms on alive nodes.

E. Analysis on Risk and Total Packets Transmitted to Base Station

Fig. 7 depicts the assessment of risk associated with the CIOO method in comparison to PSO [33], DMOSC-MHRS [32], GOA [34], SMO [35], BOA [36], COA [37], and OOA [38] for the purpose of optimal cluster head selection. It is imperative to reduce the risk rate when aiming for the optimal selection of cluster heads. In this regard, the CIOO approach consistently demonstrated the lowest level of risk when compared to the conventional schemes throughout all rounds. Mainly, round=1000, the CIOO method achieved the lowest risk level of 0.012, while PSO [33], DMOSC-MHRS [32], GOA [34], SMO [35], BOA [36], COA [37], and OOA [38] exhibited higher risk ratings.

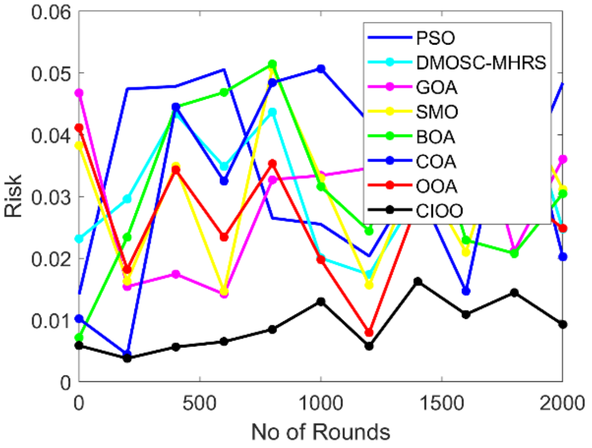


Fig. 7. Risk factor analysis of CIOO with conventional algorithms

The CIOO and conventional strategies is analyzed in terms of total packets transmitted to BS for cluster-based routing in WSN. The results of this analysis are presented in Fig. 8. Furthermore, this analysis is conducted with a network of 100 nodes, aiming to achieve the highest possible number of packets transmitted for optimal cluster-based routing. Primarily, the CIOO transmitted a larger number of packets to the BS compared to PSO [33], DMOSC-MHRS [32], GOA [34], SMO [35], BOA [36], COA [37], and OOA [38]. Hence, the CIOO approach consistently reduced risk ratings while increasing the overall number of packets sent to the BS when compared to conventional methods. In conclusion, the CIOO methodology demonstrates superior performance compared to previous approaches.

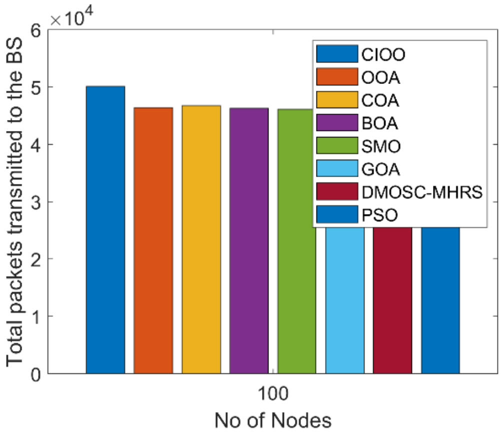


Fig. 8. Validation of CIOO and conventional schemes on total packets transmitted to base station.

F. Friedman Test Analysis

The Friedman test assessment on CIOO is compared with PSO [33], DMOSC-MHRS [32], GOA [34], SMO [35], BOA [36], COA [37] and OOA [38] for cluster-based routing in WSN is summarized in Table I. The Friedman test is a statistical hypothesis test designed for evaluating whether there exist statistically significant distinctions among various groups when analyzing correlated, non-parametric data. This test is commonly employed when dealing with multiple treatments or conditions, aiming to determine if there are overall differences in their effects. The procedure entails assigning rankings to the data within each group and subsequently assessing whether these rankings display significant variations among the groups. The presence of such variations indicates significant differences between the groups. Here, the CIOO attained the minimal value of 1, whereas the PSO [33] (5.900), DMOSC-MHRS [32] (3.500), GOA [34] (5.00), SMO [35] (6.500), BOA [36] (4.500), COA [37] (6.900) and OOA [38] (2.700), respectively.

TABLE I. ANALYSIS ON FRIEDMAN TEST

Methods	Values
PSO [33]	5.900
DMOSC-MHRS [32]	3.500
GOA [34]	5.000
SMO [35]	6.500
BOA [36]	4.500
COA [37]	6.900
OOA [38]	2.700
CIOO	1

V. CONCLUSION AND FUTURE SCOPE

This investigation proposes a new optimal cluster head selection model for WSNs based on multi constraints like delay distance security and risk. For cluster head selection, a brand new CIOO (Chimp Integrated Osprey Optimization) method has been developed. The proposed work could be evaluated with the conventional methods in terms of delay, distance, the number of alive nodes, residual energy, risk, total packets transmitted to the BS. And it is stated that the proposed CIOO method consistently outperforms against the conventional methods. These results suggest that CIOO is an effective and efficient approach for cluster-head selection in WSN, providing better network performance and energy efficiency while minimizing delay, distance and risk.

The proposed algorithm is better suited for higher-level applications where energy efficiency and the number of alive nodes are of critical concern. It may be possible to create sophisticated optimization algorithms to solve real-world problems like healthcare. Node failure detection may be an interesting security concern in the future.

REFERENCES

[1] Greeshma Arya, Ashish Bagwari and Durg Singh Chauhan, "Performance Analysis of Deep Learning-Based Routing Protocol for an Efficient Data Transmission in 5G WSN Communication", IEEE Access, volume 10, 2022, pp: 9340-9356, doi : 10.1109/ACCESS.2022.3142082.

[2] Hai-yu Zhang, "An In-depth Analysis of Uneven Clustering Techniques in Wireless Sensor Networks" International Journal of Advanced Computer Science and Applications(IJACSA), 14(3), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140381.

[3] K. B. Vikhyath and N. A. Prasad, "Combined Osprey-Chimp Optimization for Cluster Based Routing in Wireless Sensor Networks: Improved DeepMaxout for Node Energy Prediction", *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 6, pp. 12314–12319, Dec. 2023, https://doi.org/10.48084/etasr.6542.

[4] Achyutha Prasad N., Chaitra, H. V., Majula, G., Shabaz, M., Martinez-Valencia, A.B., Vikhyath, K .B.,Verma, S., & Arias-Gonzales, J. L. (2023). Delay optimization and energy balancing algorithm for improving network lifetime in fixed wireless sensor networks. Physical Communication, 58, 102038, DOI: 10.1016/j.phycom.2023.102038.

[5] Sathyaprakash B. P and Manjunath Kotari, "Dynamic Routing Using Petal Ant Colony Optimization for Mobile Ad-hoc Networks" International Journal of Advanced Computer Science and Applications(IJACSA), 14(10), 2023. http://dx.doi.org/10.14569/ IJACSA.2023.0141084.

[6] Vikhyath K B and Achyutha Prasad N (2023), Optimal Cluster Head Selection in Wireless Sensor Network via Multi-constraint Basis using Hybrid Optimization Algorithm: NMJSOA. IJEER 11(4), 1087-1096. DOI: 10.37391/ijeer.110428.

[7] Chunfen HU, Haifei ZHOU and Shiyun LV, "Clustering Based on Gray Wolf Optimization Algorithm for Internet of Things over Wireless Nodes" International Journal of Advanced Computer Science and Applications(IJACSA), 14(6), 2023. http://dx.doi.org/10.14569/ IJACSA.2023.0140637.

[8] Zongshan Wang, Hongwei Ding, Bo Li, Liyong Bao and Zhijun Yang, "An Energy Efficient Routing Protocol Based on Improved Artificial Bee Colony Algorithm for Wireless Sensor Networks", IEEE Access, volume: 8, 2020, pp: 133577-133596.

[9] Zeyu Sun, Lili Wei, Chen Xu, Tian Wang, Yalin Nie, Xiaofei Xing and Jianfeng Lu, "An Energy-Efficient Cross-Layer-Sensing Clustering Method Based on Intelligent Fog Computing in WSNs", IEEE Access, volume 7, 2019, PP:144165-144177, doi : 10.1109/ACCESS. 2019.2944858.

[10] Vikhyath K B., Brahmanand S H, Wireless sensor networks security issues and challenges: A survey. International Journal of Engineering & Technology 7(2.33), 89-94 (2018), DOI: 10.14419/ijet.v7i2.33.13861.

[11] Kale Navnath Dattatraya, K. Raghava Rao, " Hybrid based cluster head selection for maximizing network lifetime and energy efficiency in WSN", Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 3, March 2022, Pages 716-72Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 3, March 2022, Pages 716-726, doi : https://doi.org/10.1016/j.jksuci.2019.04.003.

[12] Bandi Rambabu, A. Venugopal Reddy, Sengathir Janakiraman, "Hybrid Artificial Bee Colony and Monarchy Butterfly Optimization Algorithm (HABC-MBOA)-based cluster head selection for WSNs", Journal of King Saud University – Computer and Information Sciences, volume: 34 (2022), pp: 1895–1905, doi : 10.1016/j.jksuci.2019.12.006.

[13] Indra Kumar Shah, Tanmoy Maity, Yogendra Singh Dohare, Devvrat Tyagi, Deepak Rathore and Dharmendra Singh Yadav, "ICIC: A Dual Mode Intra-Cluster and Inter-Cluster Energy Minimization Approach for Multihop WSN", IEEE Access, Volume 10, 2022, pp:70581-70594, doi: 10.1109/ACCESS.2022.3188684.

[14] Mohd Adnan, Liu Yang, Tazeem Ahmad and Yang Tao, "An Unequally Clustered Multi-Hop Routing Protocol Based on Fuzzy Logic for Wireless Sensor Networks", IEEE Access, Volume: 9, 2021, pp: 38531-38545, doi: 10.1109/ACCESS.2021.3063097.

[15] Khalid Haseeb, Naveed Islam Ahmad Almogren and Ikram Ud Din, "Intrusion Prevention Framework for Secure Routing in WSN-Based Mobile Internet of Things", IEEE Access, Volume: 7, 2019, pp: 185496-185505.

[16] I. S. Akila and R. Venkatesan, "A Fuzzy Based Energy-aware Clustering Architecture for Cooperative Communication in WSN," The Computer Journal, vol. 59, no. 10, pp. 1551-1562, Oct. 2016.

[17] Guangyue Kou and Guoheng Wei, “Hybrid Particle Swarm Optimization-based Modeling of Wireless Sensor Network Coverage Optimization” International Journal of Advanced Computer Science and Applications(IJACSA), 14(5), 2023. http://dx.doi.org/10.14569/ IJACSA.2023.01405102.

[18] S. H. Kang and T. Nguyen, "Distance Based Thresholds for Cluster Head Selection in Wireless Sensor Networks," IEEE Communications Letters, vol. 16, no. 9, pp. 1396-1399, September 2012.

[19] A. R. Ansari and S. Cho, "CHESS-PC: Cluster-Head Selection Scheme with Power Control for Public Safety Networks," IEEE Access, vol. 6, pp. 51640-51646, 2018.

[20] D. Jia, H. Zhu, S. Zou and P. Hu, "Dynamic Cluster Head Selection Method for Wireless Sensor Network," IEEE Sensors Journal, vol. 16, no. 8, pp. 2746-2754, April15, 2016.

[21] Payal Khurana Batra, Krishna Kant," LEACH-MAC: a new cluster head selection algorithm for Wireless Sensor Networks", Wireless Networks, vol.22, no.1, pp 49–60, January 2016.

[22] Zhu Xiaorong, Shen Lianfeng," Near optimal cluster-head selection for wireless sensor networks," Journal of Electronics (China), vol.24, no.6, pp 721–725, November 2007.

[23] Devesh Pratap Singh, R. H. Goudar, Bhasker Pant, Sreenivasa Rao," Cluster head selection by randomness with data recovery in WSN", CSI Transactions on ICT, vol.2, no.2, pp 97–107, June 2014.

[24] Amit Sarkar, T. Senthil Murugan," Cluster head selection for energy efficient and delay-less routing in wireless sensor network", Wireless Networks, pp 1–18, 15 July 2017.

[25] Buddha Singh, Daya Krishan Lobiyal," A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks", Human-centric Computing and Information Sciences, 2:13, December 2012.

[26] Pawan Singh Mehra, Mohammad Najmud Doja, Bashir Alam," Fuzzy based enhanced cluster head selection (FBECS) for WSN", Journal of King Saud University – Science, Available online 27 April 2018.

[27] A. Amuthan, A. Arulmurugan," Semi-Markov inspired hybrid trust prediction scheme for prolonging lifetime through reliable cluster head selection in WSNs", Journal of King Saud University - Computer and Information Sciences, Available online 17 July 2018.

[28] Shilpa Mahajan, Jyoteesh Malhotra, Sandeep Sharma," An energy balanced QoS based cluster head selection strategy for WSN", Egyptian Informatics Journal, vol.15, no.3, pp.189-199, November 2014.

[29] Bilal Muhammad Khan, Rabia Bilal, Rupert Young," Fuzzy-TOPSIS based Cluster Head selection in mobile wireless sensor networks", Journal of Electrical Systems and Information Technology, Available online, 4 January 2017.

[30] Wei Liu, Peng Zou, Dingguo Jiang, Xiufeng Quan, Huichao Dai, "Zoning of reservoir water temperature field based on K-means clustering algorithm", Journal of Hydrology: Regional Studies, Volume 44 (2022), doi: 10.1016/j.ejrh.2022.101239.

[31] Achyut Shankar, Natarajan Jaisankar, Mohammad S. Khan, Rizwan Patan, Balusamy Balamurugan, "Hybrid model for security-aware cluster head selection in wireless sensor networks", IET Wireless Sensor Systems, ISSN 2043-6386, 2018, doi: 10.1049/iet-wss.2018.5008.

[32] Fatma S. Alrayes, Jaber S. Alzahrani, Khalid A. Alissa, Abdullah Alharbi, Hussain Alshahrani, Mohamed Ahmed Elfaki, Ayman Yafoz, Abdullah Mohamed and Anwer Mustafa Hilal, "Dwarf Mongoose Optimization-Based Secure Clustering with Routing Technique in Internet of Drones", Drones, Vol.6, 2022.

[33] Pushpalatha A, Mahima.R, Kiruthik Ruba K S, Mohanraj E, Rajaram P, Ramesh S, "Optimized Data Routing using PSO in WSN", International Journal of Advanced Science and Technology, vol.29, 2020.

[34] Meraihi, Yassine, Asma Benmessaoud Gabis, Seyedali Mirjalili, and Amar Ramdane-Cherif. "Grasshopper optimization algorithm: theory, variants, and applications." Ieee Access 9 (2021): 50001-50024.

[35] Sharma, Harish, Garima Hazrati, and Jagdish Chand Bansal. "Spider monkey optimization algorithm." Evolutionary and swarm intelligence algorithms (2019): 43-59.

[36] Arora, Sankalap, and Satvir Singh. "Butterfly optimization algorithm: a novel approach for global optimization." Soft Computing 23 (2019): 715-734.

[37] M. Khishe, M. R. Mosavi, "Chimp Optimization Algorithm", Expert Systems with Applications, February 2020, doi: 10.1016/j.eswa.2020.113338.

[38] Mohammad Dehghani and Pavel Trojovský, "Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems", Frontiers in Mechanical Engineering, 2023, doi: 10.3389/fmech.2022.1126450.

Deep Learning-based Pothole Detection for Intelligent Transportation: A YOLOv5 Approach

Qian Li*, Yanjuan Shi, Qing Liu, Gang Liu
College of Vehicle and Transportation Engineering, Henan Institute of Technology, Xinxiang 453000, Henan, China

***Abstract*—Pothole detection plays a crucial role in intelligent transportation systems, ensuring road safety and efficient infrastructure management. Extensive research in the literature has explored various methods for pothole detection. Among these approaches, deep learning-based methods have emerged as highly accurate alternatives, surpassing other techniques. The widespread adoption of deep learning in pothole detection can be justified by its ability to learn discriminative features, leading to improved detection performance automatically. Nevertheless, the present research challenge lies in achieving high accuracy rates while maintaining non-destructiveness and real-time processing. In this study, we propose a deep learning model according to the YOLOv5 architecture to address this challenge. Our method includes generating a custom dataset and conducting training, validation, and testing processes. Experimental outcomes and performance evaluations show the suggested method's efficacy, showcasing its accurate detection capabilities.**

Keywords—Pothole detection; deep learning; intelligent transportation systems; YOLOv5

I. Introduction

Modern cities are witnessing a rapid transformation with the integration of advanced technologies, giving rise to the concept of smart cities [1]. These cities leverage cutting-edge infrastructure, digital connectivity, and intelligent systems to enhance the quality of life for their residents [2, 3]. The deployment of smart city facilities enables efficient management of resources, optimized transportation systems, and improved public services [4, 5]. However, amidst these advancements, the issue of deteriorating road conditions, particularly potholes, remains a persistent challenge. Potholes not only pose a threat to the safety of road users but also result in increased maintenance costs and traffic disruptions [6]. Therefore, the development of effective pothole detection methods has become crucial in modern cities.

Detecting and monitoring potholes in real-time is of paramount importance for maintaining the infrastructure and ensuring the safety of road users. Traditional manual inspection methods are time-consuming, expensive, and often inefficient for large-scale monitoring [7]. As a result, there is a growing interest in leveraging automated technologies for pothole detection. By employing advanced sensors, data analytics, and machine learning algorithms, cities can detect and address potholes promptly, thereby minimizing the associated risks and reducing repair costs.

In recent years, several technologies and methodologies have been proposed for pothole detection. Vision-based methods have gained a lot of attention because of their non-destructive nature and real-time applicability [4, 8, 9]. Vision-based techniques utilize cameras and image processing algorithms to analyze road surface images and identify potential potholes [10, 11]. These methods offer advantages such as cost-effectiveness, simplicity, and compatibility with existing surveillance infrastructure. Furthermore, with recent advances in deep learning techniques, vision-based pothole detection has witnessed substantial improvements in accuracy and robustness. Fig. 1 demonstrates a scheme of vision-based pothole detection system.

Previous studies have explored various vision-based and deep learning-based methods for pothole detection. Deep learning, in particular, has attracted significant interest from researchers due to its ability to learn discriminative features from large-scale datasets automatically. Convolutional neural networks (CNNs), one type of deep learning model, have displayed remarkable performance in a variety of computer vision tasks [13-15]. Consequently, researchers have investigated deep learning-based approaches for pothole detection, aiming to overcome the limitations of traditional methods and achieve higher accuracy.

However, despite the advancements in deep learning-based pothole detection, there are still research challenges and limitations that need to be addressed. Achieving high accuracy, non-destructiveness, and real-time processing capabilities remains demanding. To overcome these challenges, further studies are necessary to develop innovative approaches that meet the requirements of modern cities and their infrastructure management systems.

This investigation suggests a deep learning-based method for pothole detection using video analysis. By adopting a deep learning approach, we aim to tackle the aforementioned research challenges as well as satisfy the needs of high accuracy, non-destructiveness, and real-time processing. To train our model, we generate a custom dataset specifically designed for the pothole detection challenge. This dataset encompasses diverse road conditions and a wide range of pothole instances. Through training, validation, and testing processes, we demonstrate the effectiveness of our suggested method.

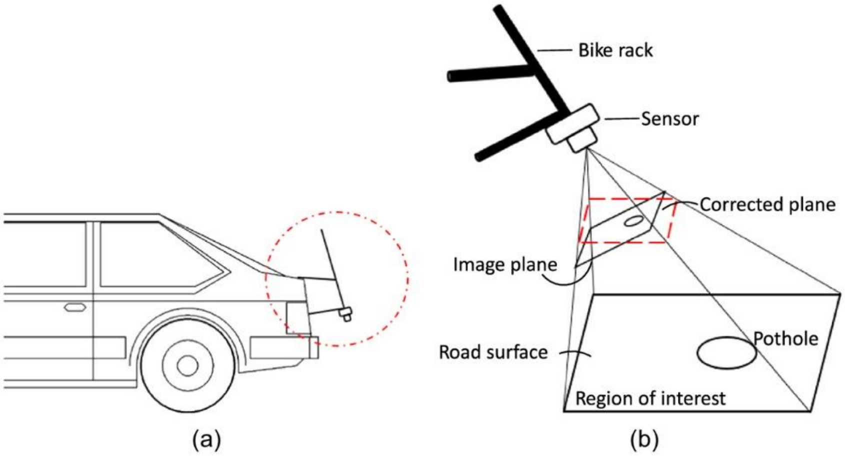


Fig. 1. A scheme of vision-based pothole detection [12].

The significance of this study lies in its pioneering approach to pothole detection through video analysis and the implementation of cutting-edge deep learning techniques, particularly the YOLOv5 algorithm. By addressing the challenges associated with pothole detection, such as high accuracy requirements, non-destructiveness, and real-time processing, our research contributes to the advancement of intelligent transportation systems.

The contributions of this research can be summarized as follows. Firstly, we generated a custom dataset tailored for the pothole detection challenge, providing a comprehensive benchmark for future studies. Secondly, we propose an efficient deep learning method that demonstrates superior performance in pothole detection compared to existing approaches. Lastly, we carry out extensive experiments and performance evaluations to verify the efficacy and viability of our approach, thereby contributing to the body of knowledge in this field.

The rest of this paper is as follows, Section II review previous studies. Section III discusses the methodology. Section IV presents experimental results. Finally, conclusion presents in Section V.

II. Related Works

Ping et al. [16] developed a deep learning-based method for street pothole detection. The proposed approach utilizes video analysis to identify and locate potholes in real-time accurately. The method employs a custom dataset specifically designed for the pothole detection challenge, ensuring diverse road conditions and pothole instances. Through extensive training, validation, and testing processes, the deep learning model demonstrates superior performance compared to existing approaches. However, the study acknowledges the limitation of demanding high accuracy, non-destructiveness, and real-time requirements, which pose research challenges. Further investigation is needed to address these limitations and develop innovative solutions to meet the needs of modern cities and infrastructure management systems. As a result, this research contributes by providing a comprehensive benchmark dataset, proposing an efficient deep-learning method, and conducting extensive performance evaluations for street pothole detection.

Pandey et al. [2] focused on pothole detection of critical road infrastructure using convolutional neural networks (CNNs). The method employs CNNs to analyze road surface images and accurately identify potholes. The study demonstrates the effectiveness of the proposed approach through extensive experiments and evaluations. However, the limitation of the method lies in its dependence on high-quality images and the need for sufficient training data. Further research is required to address these limitations and enhance the method's performance for real-world applications. In summary, this research contributes by utilizing CNNs for pothole detection, highlighting the method's potential, and identifying areas for future improvement.

Ahmed in study [17] presented a smart pothole detection method based on deep learning using dilated convolution. The proposed approach leverages dilated convolutional neural networks (DCNNs) to analyze road surface images and accurately detect potholes. The method takes advantage of the dilation technique to capture both local and global contextual information, enhancing the detection performance. Extensive experiments are conducted to validate the effectiveness of the proposed approach. However, the limitation of the method lies in its reliance on high-quality and well-segmented images, which may pose challenges in real-world scenarios with varying lighting conditions and road surface textures. Further research is needed to address these limitations and improve the method's robustness and generalizability. Therefore, this research contributes by introducing a deep learning-based approach with dilated convolution for smart pothole detection and highlighting the potential of this technique in enhancing road infrastructure management systems.

In research [18], a real-time pothole detection method was proposed using the YOLOv5 algorithm, aiming to enhance intelligent transportation systems. The approach utilizes the YOLOv5 architecture, which is a state-of-the-art object detection algorithm, to detect potholes in road images accurately. The method focuses on achieving real-time processing capabilities, enabling prompt identification and response to potholes. The study validates the feasibility of the approach through experiments and evaluations. However, one limitation of the method is its reliance on high-quality and well-annotated training data, which may pose challenges in real-world scenarios with diverse road conditions and variations in pothole appearances. Further research is necessary to address these limitations and improve the method's performance in challenging environments.

III. Methodology

In our study, a YOLOv5 model was generated for pothole detection on a custom dataset. The architecture of YOLOv5 is shown in Fig. 2 [19]. The dataset was annotated, meaning that each image was manually labeled to indicate the location and extent of potholes. This annotation process involved carefully marking the boundaries of potholes to create bounding box annotations. Additionally, class labels were assigned to differentiate potholes from other objects in the images.

To enhance the dataset and improve model performance, data augmentation techniques were applied. These techniques involved transforming the original images to create additional variations. In our study, data augmentation was performed to cover a broader range of scenarios and appearance variations. Various transformations, such as rotation, scaling, flipping, and changes in lighting conditions, were applied to the images. By augmenting the dataset, we aimed to increase its diversity and enable the model to generalize better and handle different real-world scenarios.

Subsequently, the dataset was split into three subsets: training, validation, and testing sets. The split was performed using an 80%, 13%, and 7% ratio, respectively. The training set, comprising 80% of the dataset, was utilized to train the YOLOv5 model. During the training phase, the model learns from the annotated images, optimizing its parameters to improve pothole detection accuracy. The validation set, accounting for 13% of the dataset, was utilized to appraise the model's performance during training. It helps in monitoring the model's progress, identifying potential overfitting, and making necessary adjustments to improve accuracy. Lastly, the testing set, consisting of 7% of the dataset, was utilized to assess the final performance of the trained YOLOv5 model. It provides an unbiased evaluation of the model's ability to detect potholes in unseen data.

After preparing the annotated and augmented dataset, we split it into three subsets: training, validation, and testing sets. The split was based on a specific ratio, with 80% of the dataset allocated to the training set, 13% to the validation set, and 7% to the testing set.

The training set, which accounts for the majority (80%) of the dataset, is utilized to train the YOLOv5 model. During training, the model learns from the annotated images, optimizing its parameters to improve pothole detection accuracy. By exposing the model to a diverse range of pothole images and their corresponding annotations, it becomes capable of recognizing and localizing potholes effectively. Training typically involves iterating through multiple epochs, gradually refining the model's performance.

The validation set, comprising 13% of the dataset, is utilized to appraise the performance of the model during training. It serves as a means to monitor the model's progress and assess its generalization capabilities. The model's performance metrics, such as detection accuracy and loss values, are measured on this set. The validation set aids in identifying potential issues like overfitting, where the model becomes overly specialized to the training data but fails to generalize well to new, unseen examples. If overfitting is observed, adjustments can be made to the model architecture, regularization techniques, or hyperparameters to improve accuracy and generalization.

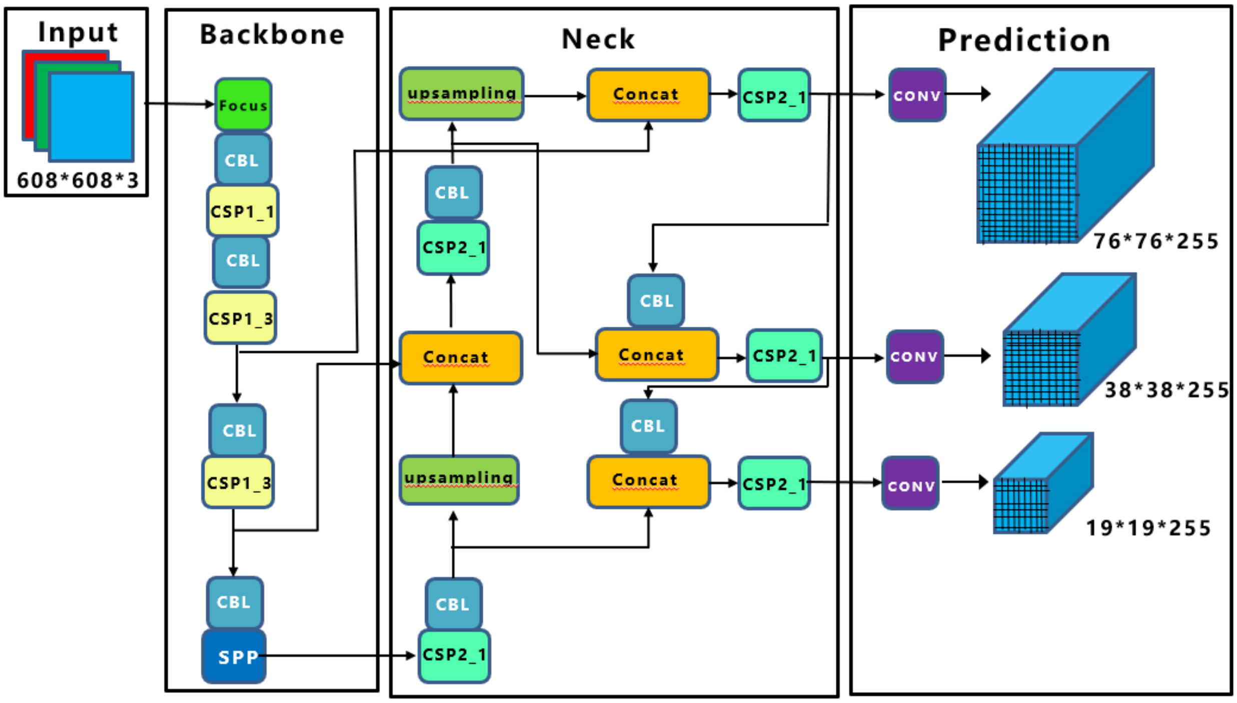


Fig. 2. The architecture of YOLOv5 [19].

Lastly, the testing set, which consists of 7% of the dataset, is reserved for the final evaluation of the trained YOLOv5 model. This set serves as an unbiased benchmark to assess the model's performance on unseen data. By applying the model to the testing set, we can measure its effectiveness in detecting potholes and evaluate its overall recall, accuracy, precision, and other relevant performance metrics. This evaluation helps to validate the model's capabilities as well as determine its readiness for deployment in real-world scenarios.

IV. Experimental Results

In this section, the YOLOv5 algorithm is compared to other algorithms. This comparison is presented to prove the YOLOv5 model enables to present better performance compared to another algorithm. Different versions of YOLOv5 and YOLO-Z models are compared in terms of the Mean Average Precision (MaP) metric. This comparison is inspired by the study in [20]. The comparison of the YOLOv5 and YOLO-Z versions is illustrated in Fig. 3.

Performance comparison between the YOLOv5 and YOLO-Z families of models, plotting mAP (top) is analysed. The superior average performance of YOLOv5 is achieved, while performance is stable and very close to 1 [20]. Performance comparison between the YOLOv5 and YOLO-Z families of models was conducted, with the mean average precision (mAP) plotted for evaluation. The results clearly indicate the superior average performance achieved by YOLOv5 compared to the YOLO-Z models. The mAP values for YOLOv5 remained stable and consistently close to 1, indicating a high level of accuracy and reliability in pothole detection. On the other hand, the YOLO-Z models exhibited slightly lower mAP scores, suggesting a relatively lower performance in terms of pothole detection accuracy.

The superior average performance of YOLOv5 can be attributed to several factors. Firstly, YOLOv5 benefits from architectural improvements and optimizations compared to the YOLO-Z models. These enhancements allow YOLOv5 to effectively capture and analyze visual features relevant to pothole detection, resulting in higher precision and recall rates. Secondly, YOLOv5 incorporates advanced training strategies and data augmentation techniques, enabling the model to generalize well to diverse road conditions and pothole appearances. This robustness contributes to its stable and consistently high performance across various scenarios.

A. Results of Yolov5 Models in Different Architectures

The performance of YOLOv5 models on Google Colab is well-established, but their performance on mobile devices remains uncertain. While the model's accuracy is unaffected by the execution platform, factors such as available resources and architecture can impact its performance and inference time. To assess the viability of using YOLOv5 models in real-time on mobile devices that inspired from study [21], several experiments are conducted using an iPhone 12 equipped with different system-on-a-chip (SoC) components. These components included an Apple Neural Engine (ANE), a graphical processing unit (GPU), and a central processing unit (CPU).

As shown in Fig. 4, we observed consistent patterns similar to those observed in the previous experiment. It is worth noting that the YOLOv5s model, which is the smallest in terms of size, exhibited the shortest inference time. However, as the complexity of the model increased, its inference speed decreased proportionally. Among the architectures we examined, it became clear that the ANE architecture demonstrated the fastest inference. This characteristic makes it particularly well-suited for running YOLOv5 models on the iPhone 12.

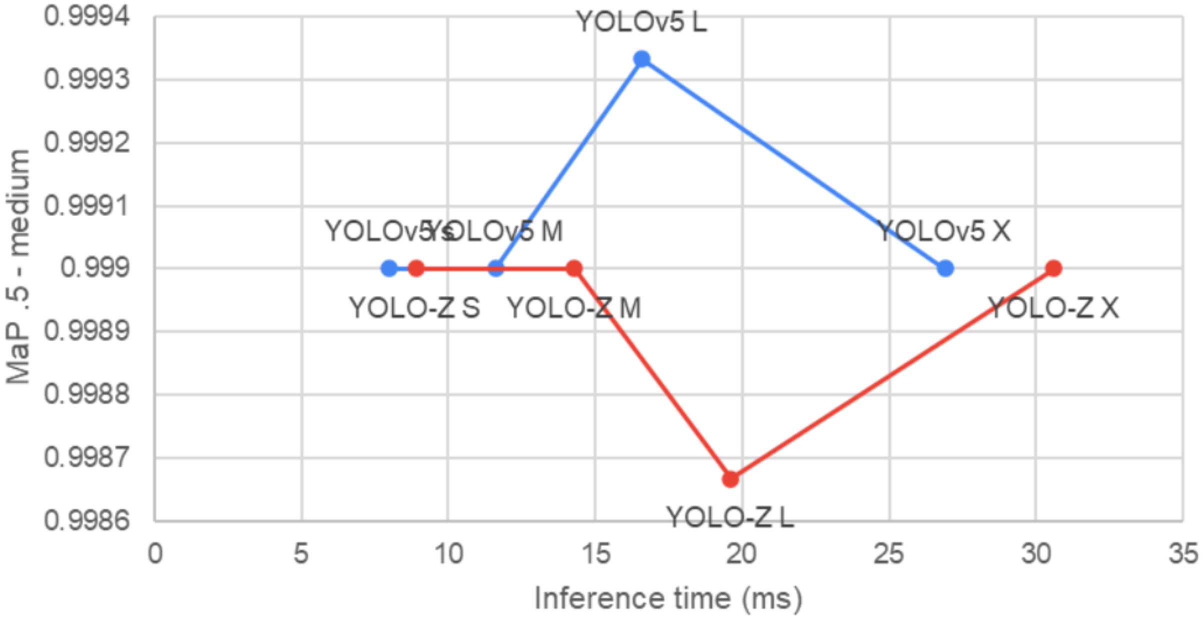


Fig. 3. Performance comparison between the YOLOv5 and YOLO-Z families of models [20].

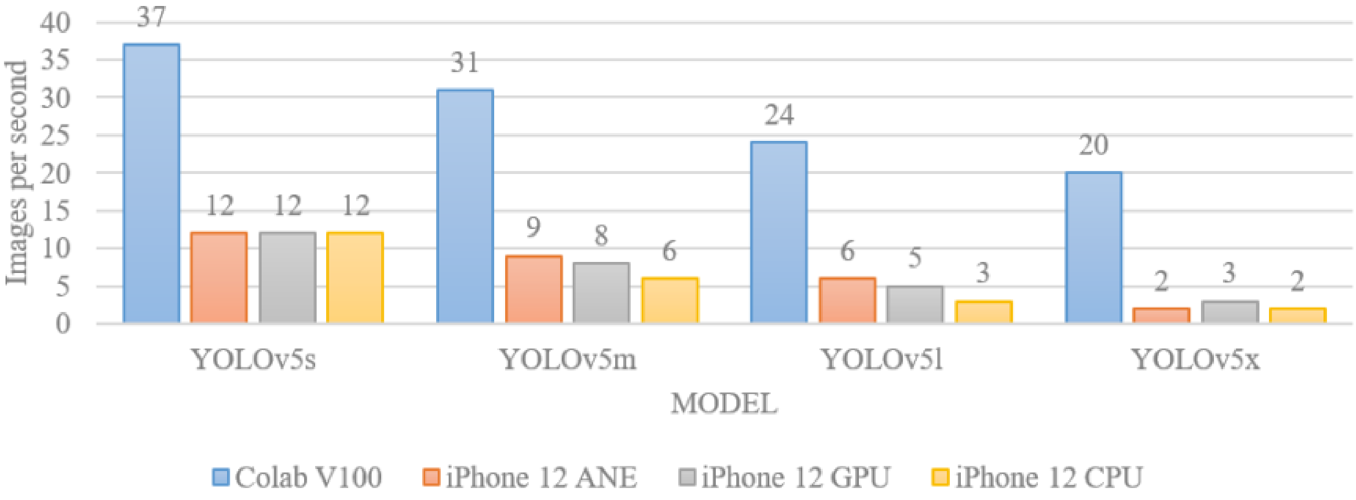


Fig. 4. Comparison of different architectures of YOLOv5 models [21].

Through the comparison of the different models, we encountered similar trends as in the previous experiment. The YOLOv5s model, due to its smaller size, required the least amount of time for inference. Conversely, as the model's complexity grew, its inference speed decreased accordingly. Notably, when considering the various architectures, the ANE architecture stood out for its superior performance in terms of inference speed. Consequently, for optimal execution of YOLOv5 models on the iPhone 12, the ANE architecture emerged as the most appropriate choice, as it consistently delivered the fastest inference results.

B. *Results of our Experiments*

In this study, the proposed YOLOv5 model, generated through our training and validation processes, is implemented and tested on various image sets. The experimental results, illustrating the performance of our model, are presented in Fig. 5. The figure provides a visual representation of the effectiveness and accuracy of our YOLOv5-based approach in detecting potholes. Through these experiments, we demonstrate the practicality and robustness of our model, highlighting its potential for real-world applications in pothole detection tasks.

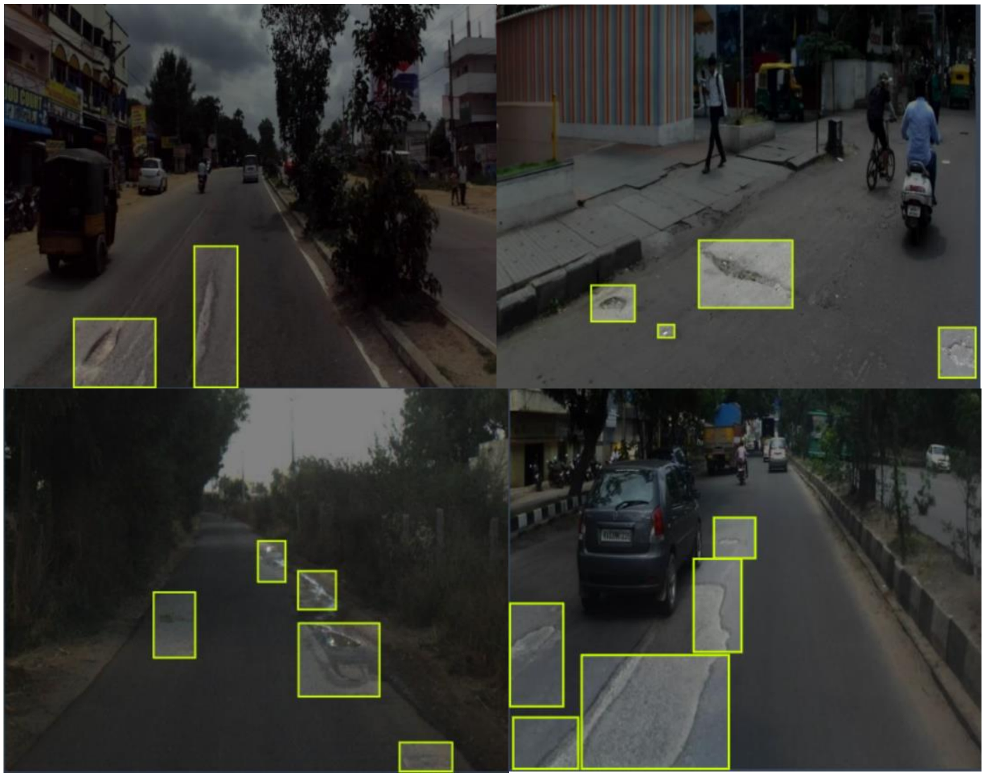


Fig. 5. Sample of experimental result.

C. *Performance Evaluation*

Performance evaluation of a generated YOLOv5 model is typically assessed by analyzing the training and validation losses. The training loss measures how well the model is learning during the training phase, while the validation loss evaluates its performance on unseen data. By examining these two metrics, one can gain insights into the model's ability to generalize as well as make accurate predictions. During the training process, the YOLOv5 model's training loss is monitored and analyzed. The training loss indicates the disparity between the forecasted bounding boxes and the ground truth labels for the training images. A lower training loss suggests that the model is effectively learning to detect objects and minimizing the error between its predictions and the actual objects present in the images. However, it is important to strike a balance between reducing the training loss as well as avoiding overfitting, where the model becomes too specialized to the training data and fails to generalize well to new data. Fig. 6 shows training/loss graphs of the generated model.

The validation loss is evaluated using a separate set of images that the model has not seen during training. This metric provides an estimate of how well the model is performing on new, unseen data. A low validation loss indicates that the model is generalizing well and making accurate predictions on unfamiliar images. If the validation loss is significantly higher than the training loss, it suggests that the model may be overfitting to the training data, highlighting the need for regularization techniques such as dropout or data augmentation to improve generalization. By closely monitoring the training and validation losses, researchers and practitioners can iteratively refine the YOLOv5 model to achieve better performance and enhance its object detection capabilities. Fig. 7 shows validation/loss graphs of the generated model.

In YOLOv5, the graphs involving "train/box_loss," "train/obj_loss," and "train/cls_loss" provide insights into the training process and help us achieve accuracy in the pothole detection model. Training a YOLOv5 model for pothole detection involves monitoring the "train/box_loss," "train/obj_loss," and "train/cls_loss" graphs. These graphs provide insights into the model's performance during training and are crucial for achieving accuracy.

A lower "train/box_loss" indicates that the model is accurately localizing potholes by predicting precise bounding box coordinates. Reducing the "train/obj_loss" demonstrates that the model is effectively distinguishing potholes from other objects or background areas. Minimizing the "train/cls_loss" suggests that the model is correctly classifying potholes. To achieve accuracy, data augmentation techniques, such as applying transformations to training data, help the model generalize better.

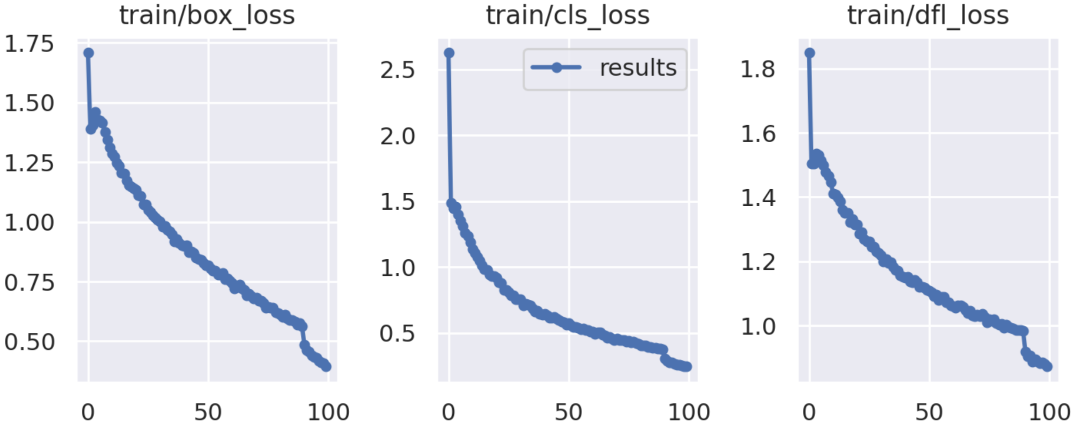


Fig. 6. Training/losses graphs of the generated model.

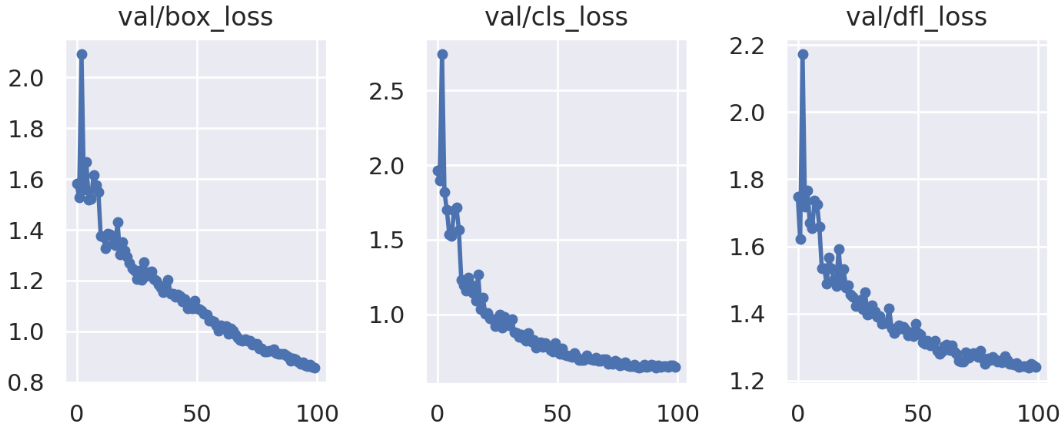


Fig. 7. Validation/loss graphs of the generated model.

Optimizing hyperparameters specific to YOLOv5, such as learning rate and weight decay, is crucial. Monitoring and analyzing the loss graphs during training is vital, allowing adjustments to be made if any loss value becomes stagnant or starts increasing. Through an iterative process of training, evaluation, and adjustment, the YOLOv5 model can learn and refine its pothole detection capabilities, ultimately achieving accuracy in identifying and classifying potholes.

In YOLOv5, the validation graphs involving "val/box_loss," "val/obj_loss," and "val/cls_loss" provide insights into the performance of the pothole detection model during the validation phase. These graphs help us evaluate and improve the accuracy of the model. In the following, we discuss each of these validation graphs and how they contribute to achieving an accurate model:

"Val/box_loss" graph: This graph represents the box regression loss during validation. Box regression loss measures the discrepancy between predicted bounding box coordinates and the ground truth coordinates of the potholes. A lower box loss indicates that the model is accurately localizing the potholes' positions. To achieve an accurate model, you would monitor the "val/box_loss" graph and aim to minimize it over the training iterations. Decreasing box loss signifies that the model is improving its ability to predict the bounding boxes around potholes precisely.

"Val/obj_loss" graph: This graph depicts the objectness loss during validation. Objectness loss measures the confidence of the model in detecting whether a pothole exists within a bounding box. It represents the ability of the model to discriminate between potholes and non-pothole regions accurately. To achieve accuracy, you would aim to reduce the "val/obj_loss" value. Lower objectness loss shows that the model is more proficient at distinguishing potholes from other objects or background areas.

"Val/cls_loss" graph: This graph represents the classification loss during validation. Classification loss measures the accuracy of the model in assigning the correct class label (e.g., "pothole") to the detected objects. To achieve accuracy, you would strive to minimize the "val/cls_loss" value. A lower classification loss shows that the model is correctly identifying potholes and effectively distinguishing them from other object classes.

In our generated model, to achieve an accurate model, you would typically iterate on the training process, adjusting various parameters and monitoring the validation graphs. The objective is to see a gradual reduction in losses over time, demonstrating that the model is learning and improving its ability to detect and classify potholes accurately.

As results, the application of deep learning is pivotal in overcoming the inherent challenges associated with pothole detection, with our primary objectives being the attainment of elevated accuracy, non-destructiveness, and real-time processing capabilities. A critical aspect of our approach involves the development of a meticulously curated dataset, tailored specifically for training our model. This dataset encompasses diverse road conditions and various instances of potholes, ensuring the robustness and adaptability of our detection system. Through a comprehensive evaluation process, involving stringent validation, training, and testing protocols, we establish the effectiveness of our proposed method. Leveraging the YOLOv5 algorithm, our model not only refines the precision and stability of pothole detection but also contributes significantly to the overall efficiency of intelligent transportation systems. The generated model using YOLOv5 advancements not only enhances the accuracy and stability of pothole detection but also contributes to the overall efficiency of intelligent transportation systems. The results emphasize the value of adopting state-of-the-art models like YOLOv5 for real-time pothole detection tasks, ensuring optimal performance and facilitating proactive maintenance and repair of road infrastructure.

V. Conclusion

In this research, we present a novel approach for detecting potholes using video analysis and deep learning techniques. By utilizing deep learning, our aim is to address the aforementioned research challenges and meet the requirements of achieving high accuracy, non-destructiveness, and real-time processing. To train our model effectively, we have created a custom dataset specifically tailored for the pothole detection task, which covers a wide range of road conditions and various instances of potholes. Through rigorous validation, training, and testing procedures, we demonstrate the efficacy of our suggested method. The developed model, employing the YOLOv5 algorithm, not only enhances the precision and stability of pothole detection but also contributes to the overall efficiency of intelligent transportation systems. These findings underscore the significance of adopting cutting-edge models like YOLOv5 for real-time pothole detection applications. This approach ensures optimal performance and facilitates proactive maintenance and repair of road infrastructure, promoting safer and more efficient transportation networks. For future research in the field of pothole detection, the integration of additional sensor modalities, such as LiDAR or infrared imaging, alongside vision-based methods can be explored to enhance the accuracy and robustness of pothole detection systems. Moreover, investigates the use of transfer learning techniques to leverage pre-trained deep learning models on large-scale datasets from related tasks, enabling efficient and effective pothole detection even with limited training data. Furthermore, the number of experiments can be increased to obtain more relevant and accurate results that can b e investigated for future works.

Acknowledgment

Key scientific research project of colleges and universities in Henan Province (No. 22B46003); Henan Province Key R&D and Promotion Special (Science and Technology Research) Project (No. 212102210046).

References

[1] Gaikwad, M.A., et al., Road Condition Improvement in Smart Cities Using IoT. 2020.

[2] Pandey, A.K., et al., Convolution neural networks for pothole detection of critical road infrastructure. Computers and Electrical Engineering, 2022. 99: p. 107725.

[3] Chen, D., et al., Real-Time Road Pothole Mapping Based on Vibration Analysis in Smart City. IEEE Journal of selected topics in applied Earth observations and remote sensing, 2022. 15: p. 6972-6984.

[4] Ma, N., et al., Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms. Transportation safety and Environment, 2022. 4(4): p. tdac026.

[5] Bhatt, A.K. and S. Biswas, AI Enabled Road Health Monitoring System for Smart Cities. 2022, EasyChair.

[6] Muhamad, N.J.A., et al. Machine Learning Combined with Thresholding-A Blended Approach to Potholes Detection. in 2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC). 2022. IEEE.

[7] Thompson, E.M., et al., SHREC 2022: Pothole and crack detection in the road pavement using images and RGB-D data. Computers & Graphics, 2022. 107: p. 161-171.

[8] Chitale, P.A., et al. Pothole detection and dimension estimation system using deep learning (yolo) and image processing. in 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ). 2020. IEEE.

[9] Mei Choo Ang, A.A., Kok Weng Ng, Elankovan Sundararajan, Marzieh Mogharrebi, Teck Loon Lim, Multi-core Frameworks Investigation on A Real-Time Object Tracking Application. Journal of Theoretical & Applied Information Technology, 2014.

[10] Dhiman, A. and R. Klette, Pothole detection using computer vision and learning. IEEE Transactions on Intelligent Transportation Systems, 2019. 21(8): p. 3536-3550.

[11] Park, S.-S., V.-T. Tran, and D.-E. Lee, Application of various yolo models for computer vision-based real-time pothole detection. Applied Sciences, 2021. 11(23): p. 11229.

[12] Huang, Y.-T., et al., Deep Learning–Based Autonomous Road Condition Assessment Leveraging Inexpensive RGB and Depth Sensors and Heterogeneous Data Fusion: Pothole Detection and Quantification. Journal of Transportation Engineering, Part B: Pavements, 2023. 149(2): p. 04023010.

[13] Jakubec, M., et al., Comparison of CNN-Based Models for Pothole Detection in Real-World Adverse Conditions: Overview and Evaluation. Applied Sciences, 2023. 13(9): p. 5810.

[14] Arjapure, S. and D. Kalbande. Deep learning model for pothole detection and area computation. in 2021 International Conference on Communication information and Computing Technology (ICCICT). 2021. IEEE.

[15] Aghamohammadi, A., et al., Correction: A parallel spatiotemporal saliency and discriminative online learning method for visual target tracking in aerial videos. Plos one, 2018. 13(3): p. e0195418.

[16] Ping, P., X. Yang, and Z. Gao. A deep learning approach for street pothole detection. in 2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService). 2020. IEEE.

[17] Ahmed, K.R., Smart pothole detection using deep learning based on dilated convolution. Sensors, 2021. 21(24): p. 8406.

[18] SB, B.K., et al. Real-time Pothole Detection using YOLOv5 Algorithm: A Feasible Approach for Intelligent Transportation Systems. in 2023 Second International Conference on Electronics and Renewable Systems (ICEARS). 2023. IEEE.

[19] Huang, T., et al. Tiny Object Detection based on YOLOv5. in Proceedings of the 2022 5th International Conference on Image and Graphics Processing. 2022.

[20] Benjumea, A., et al., YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. arXiv preprint arXiv:2112.11798, 2021.

[21] Dlužnevskij, D., P. Stefanovič, and S. Ramanauskaite, Investigation of YOLOv5 Efficiency in iPhone Supported Systems. Baltic Journal of Modern Computing, 2021. 9(3).

Security and Privacy of Cloud Data Auditing Protocols: A Review, State-of-the-art, Open Issues, and Future Research Directions

Muhammad Farooq[1], Mohd Rushdi Idrus[2]*, Adi Affandi Ahmad[3], Ahmad Hanis Mohd Shabli[4], Osman Ghazali[5]
School of Computing, Universiti Utara Malaysia, Kedah, Malaysia[1, 2, 3, 4, 5]
Institute for Advanced and Smart Digital Opportunities (IASDO), Universiti Utara Malaysia, Kedah, Malaysia[2]

***Abstract*—Cloud service providers offer a trustworthy and resistant-based storage environment for on-demand cloud services to outsource clients' data. Several researchers and business entities currently adopt cloud services to store their data in remote cloud storage servers for cost-saving purposes. Cloud storage offers numerous advantages to users like scalability, low capital expenses, and data available from any place, anytime, regardless of location and device. However, as the users lose physical access and control over data, the storage service raises security and privacy issues, such as confidentiality, integrity, and availability of outsourced data. Data integrity is a primary concern for cloud users to confirm whether data integrity is intact or not. This paper presents a comprehensive review of cloud data auditing schemes and a comparative analysis of the desirable features. Furthermore, it provides advantages and disadvantages of the state-of-the-art techniques and a performance comparison regarding the communicational and computational costs of involved entities. It also highlights desirable features of different techniques, open issues, and future research trends of cloud data auditing protocols.**

Keywords—Cloud computing; proof of possession; data integrity auditing; proof of retrievability; public auditing; proof of ownership

I. Introduction

Cloud computing is a new, rapidly emerging technology paradigm [1, 2], which can resolve large-scale service issues in multiple industries, such as engineering, sciences, and e-commerce [3]. Currently, several Cloud Service Providers (CSPs), for example, Linode [4], Amazon EC2 [5], Google Cloud Platform [6], and Microsoft Azure [7], manage and distribute shared resources for cloud users [8, 9]. Cloud computing offers manageable shared resources like services, networks, servers, applications, and storage that can be accessed over the Internet and managed with minimal effort without the interaction of cloud service providers [10, 11]. Generally, the distributed shared resources of the CSPs are available to their clients based on the pay-as-you-go. The CSPs mainly offer services like Platform-as-a-Service, Software-as-a-Service, and Infrastructure-as-a-Service [12, 13], as shown in Fig. 1. The web applications we used years ago, for instance, YouTube, Facebook, Instagram, and Twitter, are examples of cloud computing services. Furthermore, Dropbox, Amazon web services, and Google applications are generally utilized for personal and business purposes to store and share information anywhere and anytime via the Internet.

The critical part of this business model is to outsource and share data for distributed computing. Cloud storage is an elastic on-demand service model that attains significant benefits; for example, it decreases storage administration burden and reduces infrastructure and software costs. Moreover, it helps to access data from various geographic locations, focusing on ease of maintenance, efficient computation [14], data storage, data archival disaster recovery, etc. [15]. Statista reported that the cloud computing business model persistently develops and is estimated to reach 379 billion US dollars by 2022, showing a tremendous increase in this service sector [16].

Even though the adoption and development of cloud computing businesses are growing every year, most corporations are not satisfied with using this business model due to some obstacles, like pricing, interoperability, vendor lock-in, reliability, and security [17]. Security is the most critical concern [17] because it can be compromised while transferring the data from one location to another by cloud administrators, dishonest CSPs, malware, or other malevolent users who can mutilate the data [18].

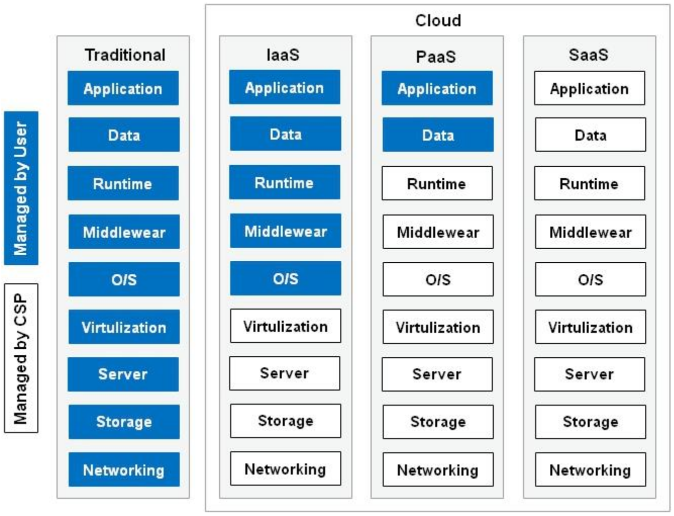


Fig. 1. Service-oriented architecture of cloud computing.

*Corresponding Author.

An insecure storage server can lead to data or privacy leakage if unauthorized users get access to data. For example, the infamous data breach incident of Microsoft's business cloud suit in which unauthorized users gained control of data [19, 20]. Statista shows that the total spending on IaaS is increasing every year [21], and with the increase in cloud demand [21], many security incidents were reported in the top CSPs [22]. The association of information audit systems defines data security as a combination of three fundamental components, i.e., confidentiality, integrity, and availability, also known as the CIA triad [23].

As contribution this paper highlight that cloud computing saves time and monitoring costs for any organization and turns technological solutions for large-scale systems into server-to-service frameworks.

A. *Background*

Cloud data storage related to IaaS is one of the critical business services offered by CSPs [24, 25]. It provides storage space for users or business organizations to host their personal or business data, as illustrated in Fig. 2. It is an ongoing trend for clients to host their data on remote storage servers. Besides the benefits, the CSPs also incorporate the critical apprehensions for the CIA security fundamentals of their client’s data. The main goal is to maintain the integrity of hosted data in the Cloud Storage Servers (CSS) [26, 27].

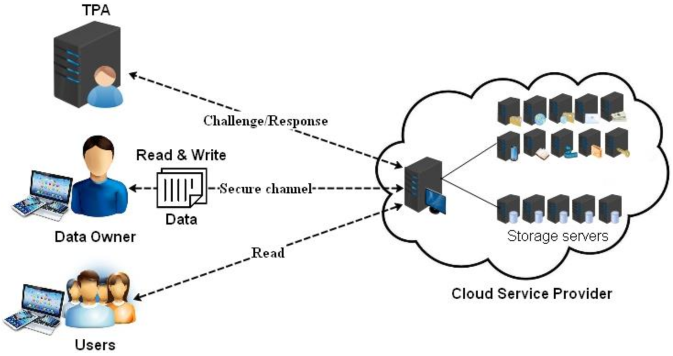


Fig. 2. The architecture of data integrity auditing protocol.

The security of remote data storage is essential because the user loses physical control of their data. One of the solutions to ensure data integrity is to utilize the fundamental cryptographic strategies based on signature and data hashing methods. However, these types of techniques need to store a local copy. Besides, it is unreasonable for the clients to retrieve all the hosted data to confirm its integrity, which incurs high communication overhead over the networks and clients and increases the communication cost. Hence, clients need auditing services for remote data storage to authenticate data integrity periodically.

Currently, researchers are focusing more on verifying the data integrity of the cloud storage servers. However, the cloud server is labelled as an adversary, and the Third-party Auditor (TPA) might see the data contents during the data auditing phase. To overcome these issues, several researchers have proposed different data security techniques. These security techniques can be mainly classified into three categories, namely Proof-of-Data Possession (PDP), Proof-of-Retrievability (PoR), and Proof-of-Ownership (PoW).

1) PDP: These techniques allow the storage server to confirm to its users that the CSPs control the hosted data with probabilistic-based assurance. However, it cannot support the recovery of data exploitation, and the data damage is irrecoverable. This causes serious concerns among cloud users, such as data loss, trust loss, financial damage, etc.

2) PoR: These techniques ensure data integrity and address the limitation of PDP techniques by supporting data exploitation recovery with Error Correction-code (ECC).

3) PoW: In these schemes, the storage servers focus on ownership of data, know which is owned by which user, and prevent downloading remote data by malicious or illegal users.

B. *Contributions*

In this article, we comprehensively studied several data integrity auditing techniques. The main contributions of this work are mentioned as follows:

1) Briefly discuss system models, basic notations, and security preliminaries used in RDA schemes.

2) Presents a comprehensive review and basic requirements of several cloud data auditing schemes.

3) Comparative analysis with desirable security features of efficient and secure data auditing protocols has been presented.

4) Provides advantages and disadvantages of the state-of-the-art data auditing schemes.

5) Evaluate the performance in terms of communication and computational cost for different entities and the data structures involved in the RDA schemes.

6) Identifies different challenges and future research trends of RDA schemes.

C. *Organizations*

Section II presents the fundamentals of Remote Data Auditing (RDA) techniques, including system models, notations, and preliminaries used to design RDA protocols. Section III presents state-of-the-art data auditing schemes with basic security requirements. The comparative analysis along desirable security parameters to ensure data integrity and overcome the privacy leakage of RDA protocols is presented in Section IV. Later, Section V presents a comparison and evaluates the performance of different protocols and data structures involved in schemes. Section VI highlights open research issues to design efficient data auditing protocols, and Section VII describes possible future research trends. Finally, we conclude the paper in Section VIII.

II. Fundamentals of RDA Techniques

This section briefly describes the system models, necessary notations, and security preliminaries used for data integrity auditing protocols.

A. System Models

Cloud data storage related to IaaS is one of the critical business services offered by CSPs [24, 25]. It provides storage space for users or business organizations to host their personal or business data, as illustrated in Fig. 2. It is an ongoing trend for clients to host their data on remote storage servers. Besides the benefits, the CSPs also incorporate the critical apprehensions for the CIA security fundamentals of their client's data. The main goal is to maintain the integrity of hosted data in the Cloud Storage Servers (CSS) [26, 27].

The outsourced data integrity auditing services generally include four entities, particularly in public data auditing (see Fig. 3); for instance, (i) Data Owner (DO): is an entity that pre-processed and outsources data to cloud storage servers; it is also capable of performing dynamic data updates through the insert, delete, and update procedures, (ii) the CSP: the entity that offers on-demand shared cloud services and responsible for storing the user data to release the storage burden of its registered users, and it is needed to react the challenges request by the auditor, (iii) Third-party Auditor (TPA): provides audit services, without downloading entire data, with high computational and communicational capabilities to the registered user who delegates his audit task, and (iv) the users: any other user or enterprise who is registered on behalf of the DO and allowed to read the outsourced data.

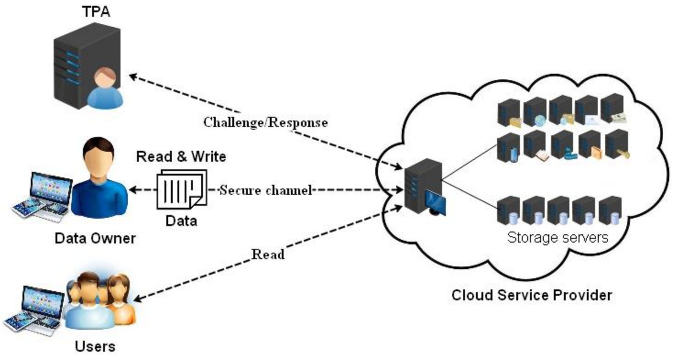


Fig. 3. Public data auditing

On the other hand, in private data auditing, a third-party auditor is not included (see Fig. 4). The outsourced data auditing technique is a challenge-response mechanism, and its comprehensive process is as follows. (i) The users initially pre-compute their data files, utilizing cryptographic, coding, or data block, and afterwards produce some metadata for all data. At that point, they transfer the data files and related metadata to the cloud storage servers. While verifying the data integrity of the hosted data, they send a challenge request to the auditor and wait for the notice of the outcome. (ii) The auditor produces an arbitrary challenge after getting the user's request and sending it to the storage servers. We regularly expect that the user approves the auditor before any activity with cloud storage servers. (iii) after getting the challenge request, the storage server produces corresponding evidence identified with the challenge and responds to the auditor. (iv) to verify the data integrity, the auditor confirms the receiving evidence. If the authentication fails, the auditor responds with a “rejected” notification to the user. Otherwise, the auditor responds “success” notice to the user, which means the outsourced data is secure.

B. Notations and Preliminaries

Homomorphic Verifiable Tag (HVT) is the foundation of current data integrity verification techniques. It can aggregate different data blocks into one value and save substantial communication costs. For enhancing audit phase proficiency, the signature mechanism is integrated with HVT to create a homomorphic signature method to verify the integrity of big data outsourced in existing data auditing techniques. As per the review of current RDA techniques, most of the data storage auditing protocols are constructed using Message Authentication Code (MAC) [28], RSA, BLS, Homomorphic Linear-authenticator (HLA) [18, 29], Identity (ID), and Certificate-less (CL) homomorphic methods. Cryptographic systems utilize these algorithms to build security primitives and cryptographic groups. Cryptographic operations are the security primitives to create the audit process, including key, tag, challenge, and proof generation phases. Nevertheless, the TPA can learn the user's data during the audit phase, which puts the data at risk [18]. To protect the privacy and integrity of data, user-initiated key, tag, or signature generation algorithms. In contrast, the cloud server and TPA execute the challenged and proof generation algorithms using public and private key security parameters.

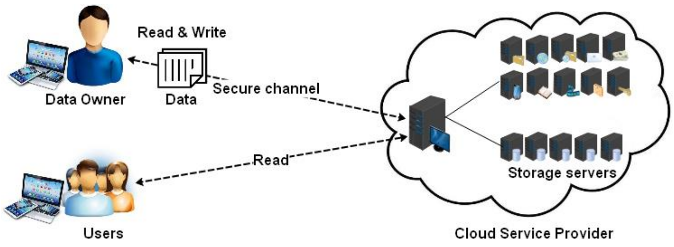


Fig. 4. Private data auditing.

A commonly utilized security parameter in a cryptosystem, denoted as n, describes the length of public and secret keys. The security parameters must be computed viably, and the implementation of the cryptographic system must be polynomial in time. When the user increases the key size, the encryption and decryption time will also be increased. It is harder for an adversary to break the system in polynomial time, generally addressed as 1n. The cryptographic procedures and keys generation for the audit reliability of the cloud storage services, and 80, 128, or 160 bits are the commonly used problem's input size security parameters [30]. The remotely stored file F is addressed as an arrangement of finite sets of memory blocks like m1, m2, m3…mn. It is essential to verify that the security parameter is not lesser than the data blocks because the user needs to encrypt the data with a relevant key [31].

The pairing function is another cryptographic method for security systems, such as let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two groups belonging to similar prime order $\mathbb{q}$, where $\mathbb{G}_1$ and $\mathbb{G}_2$ are additive and multiplicative groups, respectively. A bilinear map $e: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{GT}$ has the following properties:

- Non-degeneracy: suppose $\mathbb{P}$ is a creator of $\mathbb{G}_1$, so $e(\mathcal{P},\mathcal{P})$ is a creator of $\mathbb{G}_2$. Therefore, $e(\mathcal{P},\mathcal{P}) \neq 1$; and e proficiently computable;
- Bilinearity: $e(a\mathcal{P}, b\mathcal{Q}) = e(\mathcal{P},\mathcal{Q})^{ab}$, for all $\mathcal{P} \in \mathbb{G}_1$, $\mathcal{Q} \in \mathbb{G}_2$ and $a, b \in \mathbb{Z}_{\mathbb{q}}$, where $\mathbb{Z}_{\mathbb{q}}$ is a prime order.
- Computability: The map ‘e’ is efficiently computable.

1) Message Authentication Codes (MAC): The MAC ensures non-repudiation of the origin, validates the owner’s identity, and preserves data integrity. The hash functions need to generate the MAC codes for validation, including hash value and data block. The receiver utilizes the shared secret key known by both parties to decode and get the original message. It is easy to use MAC in the data audit phase; the user generates data blocks and their respective metadata (MACs) of data file F, uploads it to storage servers and sends the security keys to TPA. Nonetheless, this methodology needs to release the data blocks to the auditor [18]. So, cloud users can perform the integrity verification process (private audit) to prevent the TPA from validating their data. Even though the MAC authenticates the data integrity, data privacy is lost. Moreover, data integrity is at risk because attackers can modify or share the message with other users. The subsequent algorithms can be utilized in the cloud data audit phase [18]:

- generateKey($\mathcal{K}$): $\mathcal{K} \overset{\mathcal{S}}{\leftarrow} \mathcal{K}$
- generateCode: $tag \overset{\mathcal{S}}{\leftarrow} MAC_K(M)$
- verifier: $\mathcal{D} \leftarrow VF_K(M, tag)$ where $\mathcal{D} \in \{0,1\}$

The MAC uses a deterministic and state-less key generation algorithm, so it is not required a tag authentication algorithm because the receiver can create a tag by executing $MAC_K(M)$. The receiver can verify the message if the generated tag is matched to the received tag; otherwise, it fails to authenticate. Instead, using the MAC in data auditing can cause significant security issues, including the following [32]:

- It provides static data only and is unable to support dynamic data operations.
- If the user needs to update data, it requires regenerating the security keys and uploaded to the storage servers and auditor.
- The auditor should keep MAC keys because cloud users can modify their data from any geographical location.

2) Homomorphic Authentication (HA): The study in [30] was the first to introduce homomorphic linear authenticators. The HA allows the user to generate the data tag α by using the data blocks $\{M_i\}$ along with secrete keys, where $i = \{1,2,3 \dots n\}$ and store it on the cloud servers. After that, the CSP generates the data blocks $\{M_i\}$ with relative data tag by using publically accessible functions. So, the HA permits anyone to validate the output derived from the verified datasets using complex computational procedures with data tag α. Moreover, this approach also allows users to combine data stream bits of several files without revealing data contents to others.

Consider the case of supply chain management, the transactions of each department carried out for production, sales, and retail, without revealing information to other departments. The HA can be categorized as (1) partially homomorphic cryptography, which can be additive or multiplicative, and (2) fully homomorphic encryption that supports both additive and multiplicative operations [33]. The procedure to perform the storage authentication can be briefly described as follows:

- The data file φ is represented as an N vector.
- Then generated the tags τ of every data block φ.
- The user randomly generates a challenge request c and sends it to the cloud storage servers.
- The server responds as data verification proof by using:

$$\mu = \sum_j c_j, \varphi_j$$

The homomorphic provable tag has been utilized in data audit procedures. They have flexibility and Blockless authentication properties. Blockless authentication permits the cloud servers to authenticate the integrity of data deprived of computing the data and metadata of the data block. Metadata (tags) and the distinctive index are created and stored as inclusive counters for every data block. Later, the storage servers generate a proof and allow users to authenticate data integrity with linear summation of tag values [34]. The HA provides cumulative signatures, where ƞ signatures related to ƞ messages for ƞ number of users [35]; support homomorphic signature [36] and batch authentication [37].

The following four algorithms are included in the homomorphic authentication [30]:

$(\rho\kappa, s\kappa) \leftarrow (1^k)$: The data owner initiates this algorithm for the setup phase to verify data storage. It requires security parameters to compute public $\rho\kappa$ and secret $s\kappa$ keys.

$(\vec{t}, st) \leftarrow Tag_{sk}(\vec{f})$: It is also a probabilistic algorithm executed by the user to tag a file. It uses a private key $s\kappa$ and a file $\vec{f} \in [\mathbb{B}]^n$ as an input security parameter and computes the tags’ vector and state information st.

$\tau := Auth_{pk}(\vec{f}, \vec{t}, \vec{c})$: The cloud storage servers executed this deterministic method by computing a tag. It uses public-key $\rho\kappa$, a data block $\vec{f} \in [\mathbb{B}]^n$, a tag $\vec{t}$, and a challenge request $\vec{c} \in \mathbb{Z}_p^n$ as an input security parameter and computes a proof τ.

$b := Verify_{\rho\kappa}(st, \mu, \vec{c}, \tau)$: It is also a deterministic algorithm that is run by the verifier; it uses security parameters public-key $\rho\kappa$, state info st, an element $\mu \in \mathbb{N}$, challenge request $\vec{c} \in \mathbb{Z}_p^n$, a proof τ as an input, and generates a bit ‘1’ or ‘0’, accept or reject, respectively.

We determine that the private key is not required during authentication in the algorithms mentioned above. Moreover, linear data block combinations expose sufficient information to the auditor for downloading the complete file f [32, 37]. The state info belongs to $\{0,1\}^k$ which is just a security parameter resulting from tag or encode a file f algorithm. The researchers Ateniese et al. treated data file f to be n vectors. Every tagged file f could be recognized by using the state information computed in encoding algorithm [30]. The HA technique can enhance using random-masking [38] and ring-based signatures [18] schemes, particularly for public data auditing. The homomorphic authenticator-based ring signatures strategies use to share data among multiple cloud users and propose supporting data privacy and Blockless authentication. Furthermore, random masking provides privacy-preserving data auditing process.

3) RSA-based homomorphic methods: The researchers in [34] introduced a sample-based publicly audit data possession technique that integrates the RSA approach with the HVT method. The succeeding RSA-based methods mostly improved their scheme. In their scheme, some essential elements are produced similar to the RSA signature, where the public and private keys are generated as $\{N, g\}$ and $\{e, d, v\}$. The client split the data file F into n number of data blocks $F = \{x_1, x_2, \dots, x_n\}$, where $x_i \in \mathbb{Z}_q^*$. For every data block, the client creates a block tag $\sigma_i = (h.u^{x_i})(y_i)^d modulo\ N$, wherein $h: \{0,1\}^* \to R_N$ is a hash method that homogeneously maps to R_N, u is a creator of R_N, x_i denotes the i-th data block, and the value of y_i is computed by concatenating the index i with the secret value.

The value of y_i is distinct and unidentifiable for every tag. Then, the cloud storage servers store data files and their corresponding tags. Afterward, the user can confirm that the cloud server holds the data by creating a requested message for arbitrarily chosen data blocks. The storage server creates the proofs of data possession based on requested data blocks and related tags. Consequently, the user could ensure data integrity without retrieving entire data blocks. The replica storage, Curtmola et al. [39] also created a tag for data blocks similarly $\sigma_{ij} = (h.u^{x_{ij}})\left(y_{ij}\right)^d modulo\ N$, wherein x_{ij} denotes the j-th data block of i-th copy. Currently several solutions that adapted RSA method just change the hash function, such as $h(filename)$ [40] and $h(filename \parallel n \parallel i \parallel j \parallel g)$ [41].

4) BLS-based homomorphic method: This method generates a shorter signature than the RSA-based homomorphic method at the same security level; thus, most existing data auditing schemes use the BLS-based technique. The first publicly data integrity auditing technique with data dynamics features based on this method was proposed in [42]. The bilinear pairing utilized to construct the BLS technique for confirmation and signature is computed by elliptic-curve cryptography. It is an undisputable technique that verifies parties that the signature is reliable.

Moreover, the BLS could integrate with any other approach in group Diffie-Hellman assumption (GDH) $\mathbb{G}$ [43]. This procedure needs a hash-based method resulting in data space on $\mathbb{G}$. This technique could also be used in data audit schemes. Suppose $\mathbb{G} = \langle g \rangle$ prime number $\mathcal{P}$ set of the GDH, by using hash method $\mathbb{H}: \{0,1\}^* \longrightarrow \mathbb{G}$ be measured in the random oracle model. Each data block can encrypt by utilizing subsequent algorithms [43]:

- Key creation algorithm run by the cloud user by selecting an arbitrary variable, $x \overset{R}{\leftarrow} \mathbb{Z}_{\mathbb{p}}$ and generates $v \leftarrow g^x$. The public $\rho\kappa$ and secrete $s\kappa$ keys are $v \in \mathbb{G}$ and $x \in \mathbb{Z}_{\mathbb{p}}$ correspondingly.
- The tag generation algorithm $\sigma_i = (h.u^{x_i})(m_i)^x$ for each data block. The exclusion of index i supports the technique to provide data dynamics features.
- The signing algorithm utilizes a secret key and message $\mathbb{M} \in \{0,1\}^*$ computed has with $h \leftarrow H(M)$, wherein $h \in \mathbb{G}$ and $\sigma \leftarrow h^x$.
- Verify function generates $h \leftarrow H(M)$ by using a signature σ, public key $\rho\kappa$, and a data block. Therefore (g, v, h, σ) is confirmed as a legal tuple.

These techniques can further enhance supporting public data auditing and data update operations. The Merkle-binary hash-tree (MHT) is a technique to meet these objectives. Qian Wang et al. described that the MHT-based scheme hashes leave to authenticate data blocks [42]. Furthermore, to support dynamic data audits, their work extends the techniques [34, 44] to measure signatures with relative file indexes. Consequently, the previously stored data file needs to be re-computed in the data file modification process. Hence to minimize the file indexes storage overhead, the study [42] discarded the file indexes and generated tags for every data block to support dynamic data operations.

Several BLS-base techniques only change hash functions such as $h(w_i)$ [45, 46], $h(id)$ [47], $h(i)$ [48], and $h(V_i \parallel T_i)$ [49]. To provide the optimum solution, the study [31] presented a common protocol development mechanism for cloud data auditing. They divide the data file F into n number of data blocks and divide these data blocks into sectors s. Then data owner computes the tag of every block of data as $\sigma_i = \left(h.\prod_{j=1}^{s} v^{m_{ij}}\right)(m, fid \parallel i)^x$, wherein fid denotes identifier of the data block, and m_{ij} identifies the j-th sectors of relative i-th data blocks and $\{v_1, v_2, \dots, v_s\}$ are arbitrarily selected by the client. A similar technique $\sigma_i = \left(h.\prod_{j=1}^{s} v^{m_{ij}}\right)(m_i)^x$ presented by Liu et al. [50]. Later this technique was enhanced by [51] for the multi-replicas file system that also supports data update operations by generating $\sigma_i = (h.v^{m_{ij}})(m_i)^x$ block tags for cloud storage, wherein m_{ij} denotes the j-th data blocks of i-th copy. The use of partition mechanism and distinctive hash function design led to developing the different BLS-based techniques for outsourcing data auditing such as for data replication $\sigma_{ij} = (h.v^{m_{ij}})(filename \parallel n \parallel m \parallel v)^x$ [52] and $\sigma_{ij} = \left(h(m_{ij}).\prod_{j=1}^{s} v^{m_{ij}}\right)^x$ [53].

5) ID-based homomorphic methods: This technique does not support certificate features compared with RSA-based and BLS-based signatures. Wang et al. [54] used ID-based cryptography to design a PDP solution; they adapted identity aggregation signatures to develop a proof of data possession technique and demonstrated the system and security models.

The Private Key Generator (PKG) chooses the secrete key as $x \in \mathbb{Z}_p$ and generates the public key as $y = g^x \in \mathbb{G}$ in initialization process. After getting user ID, PKG generates $R = g^r$ and $\sigma = r + xh(ID, R)$ and responds to the user with the private key (R, σ), wherein $r \in \mathbb{Z}_p$ is arbitrarily chosen integer. The user computes the tag $\sigma_i = (h.v^{m_i})(i)^{\sigma}$ for a data block m_i.

The signature construction procedure is similar to BLS-based solution [48]; the only change is that the secrete key is computed by PKG. Hence, study in [55] suggested another ID signature-based PDP public data auditing technique for multi-clouds. In this method, secrete key creation is similar to the study mentioned above [54]; however, the signature structure is different because of the partitioning mechanism and cooperation among different servers. The client generates a tag as $\sigma_{ij} = \left(h.\prod_{j=1}^{s} v^{m_i j}\right)\left(N_i, GS_{l_i}, i\right)^{\sigma}$ for tuple $\left(N_i, GS_{l_i}, i\right)$ can verify the every data blocks different from each other. Nonetheless, in this technique the data security is not sufficient. It does not provide requested blocks for tagGen queries; in other words, the attacker cannot retrieve any tags of these data blocks, which is conflicting with the real capacity of the cloud storage server.

The researchers in [56] proposed an ID-based publicly provable privacy-preserving data auditing solution. Using asymmetric group key agreement suggests a substantial ID-based data auditing technique. The PKG produces the secrete key as $\sigma = (ID)^x$ and send it to the user. Then the user generates the tag $\sigma_i = \sigma^{m_i}.h(filename \parallel i)^{\eta}$ for a data block m_i, wherein $\eta \in \mathbb{Z}_p$ is arbitrarily selected by the user. The serial number integrated with the tag makes data dynamics operations impossible inconceivable.

6) Certificateless-based homomorphic methods: The Boneh–Lynn–Shacham-based technique consistently needs a reliable party to compute the user's private key; hence, data signatures can easily tamper once the party is compromised. The certificate-less cryptography settles this issue. The authors in [57] suggested the first certificate-less publicly provable data auditing scheme to validate the outsourced integrity of cloud storage. A homomorphic verifiable CL signature is constructed for the user/auditor to validate the data integrity without retrieving the entire data file, which is impossible in a conventional CL signature.

In the initial phase, the Key Generation Center (KGC) chooses to secrete the key as $X \in \mathbb{Z}_p$ and generates the public key $y = g^x \in \mathbb{G}$. The PKG receives the user's ID and generates partial-key $D = h(ID)^x$ send it to the user. Then the user chooses another partial key $X \in \mathbb{Z}_p$ to make the secrete key complete and generates public-key $y = g^x \in \mathbb{G}$. The user generates $\sigma_{ij} = (h_2.v^{m_i})(ID \parallel y \parallel id)^x.D$ tag with complete private key for a data block m_i with identifier id. Even though this method provides the solution for the private-key escrow problem with the help of KGC to generate a partial key instead of the complete key, it fails to avoid the public-key replacement attack. He et al. [58] proposed another CL technique for outsourced data integrity verification. The KGC produces the partial key in a generally unpredictable manner as follows:

- KGC sends arbitrary integer $\eta \in \mathbb{Z}_p$ and generates $\mathrm{T} = \eta.p,\ S = \eta + X.h(ID, T)\ modolus\ P$, where P denotes the multiplicative group $\mathbb{G}$.
- The KGC responds with computed partial-key (T, S), after that, the user creates a tag for the data block m_i along with id by generating $\sigma_i = (S + h_2(ID \parallel y \parallel Y).x).\left(h(id) + m_i.h(Y)\right)$ which is secure against public-key replacement and master-key retrieval attack.

The auditor can get the user's data inappropriately by solving the linear equation; this method does not ensure data privacy. Consequently, the study in [59] was introduced the CL privacy-preserving PDP technique by generating the public key with two private-key parts instead of using one part [57, 58]. The security evaluation shows that the proposed technique is verifiable and secure.

III. State-of-the-Art Data Integrity Auditng Techniques

This section presents a detailed comparative analysis of remote data integrity auditing protocols. Several security mechanisms have been projected for ensuring data integrity and overcoming privacy leakage in the literature. These schemes can be categorized according to data states: static, dynamic, privacy-preserving, single-cloud, multi-cloud, multi-owner, etc.

In study [34], the researchers proposed the first PDP protocol to address the issues of public data auditing for outsourcing the user's data on remote cloud storage servers and the authentication of data possession. They have presented two schemes: (1) the sampling-PDP method ensures strong authentication of data possession, and (2) the efficient-PDP approach supports improving proficiency with weak data possession. They used RAS-based homomorphic tags to achieve a public data auditing scheme. However, their methods only support static data, so direct data modification might raise significant security, privacy, and system design issues. The authors [60] improved their scheme and presented a dynamic version of the PDP protocol. However, their technique restricted unlimited challenge queries and supported only limited dynamic data operations, such as unable to provide block insertion.

Bowers et al. [61] introduced the HAIL protocol to provide high availability and ensure data integrity. Their protocol used the erasure codes on both single-server and multi-server layers correspondingly. It also supports the assurance for proof of data retrievability that is outsourced to remote cloud storage. Nevertheless, their protocol does not offer dynamic data and needs to store one segment of each data file locally. Moreover, it is restricted to the number of challenge queries. The authors [62] introduced the SW approach to improving the private data auditing technique presented by Shacham and Waters [63]. The server response size concerning the challenge request of the TPA is directed by the 't' set of elements along with the size of each group element in α bits, where α represented the group size in bit. They considered and improve the size of proof from

$\mathrm{O}(\mathcal{S}\alpha)$ to $\mathrm{O}(\alpha)$ in their scheme by splitting the 't' group components into two group components. It generates a longer public key compared to the technique designed by [34]. Furthermore, their approach is limited to the static nature of data and cannot ensure privacy leakage against TPA during the data auditing phase.

Stefanov et al. [64] considered the static data issue and introduced the first cloud-based PoR protocol along with a dynamic data operation named Iris. They used the MHT data structure to store and ensure outsourced data integrity. However, regardless of proficient read and write data procedure, it requires substantial bandwidth usage, server computation cost, local data space, and server input/output to verify data integrity. Like the protocol [64], Shi et al. [65] have introduced a standard MHT data structure-based publically provable technique.

Qian Wang et al. [42] improved the proof of the data possession scheme for dynamic data operations by using the MHT data structure to authenticate the block tag. They provide the solution for data dynamics and Blockless authentication. The previous studies [34, 44] describe that if the cloud server keeps a tempered data file, then the cloud server can identify this misconduct during the data authentication phase by using a data auditing algorithm with a probability of $\mathrm{O}(1)$. Moreover, this scheme cannot detect and verify minor data modifications [42]. Similarly, in [30], the researchers introduced an extended version of the data auditing protocol using ranked-based MHT, as illustrated in Fig. 5.

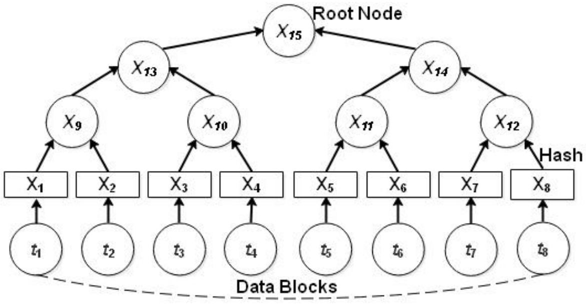


Fig. 5. The basic idea behind the MHT-based solution.

Their protocol used a signature-based approach [66] and supported authorized data auditing to avoid malicious third-party auditors. In their scheme, every node $\mathcal{N}$ should not be more than two child nodes. Every node can be denoted as $\{\mathcal{H}, r\mathcal{N}\}$, wherein $r\mathcal{N}$ representing the rank of the current node and $\mathcal{H}$ indicated hash value. Data blocks or messages m_i belongs to leaf nodes $\mathcal{LN}$, it computes hash value by $\mathcal{H}(m_i), r\mathcal{N}$, and generates the tag σ using following equation:

$$\sigma = \left(\mathcal{H}(m_a) \prod_{b=1}^{s_a} u_b m_{ab} \right)$$

where, $u_b \in \mathcal{U}, \mathcal{U} = \{\mathcal{U}_k \in \mathbb{Z}_{\mathbb{p}}\}, \mathcal{K} \in [1, s_{max}]$ the sector of data blocks s, the data file f is divided as $\{m_{ab}\}$, wherein m is the message or data block, a is the length of the current block, and b is the set of s sectors. One of the significant properties of this protocol is that an unauthorized user cannot generate a challenge for TPA without having this verification tag. Several proposed protocols cannot ensure data privacy against third-party auditors [32, 34, 42, 44]. The researchers already reported that the auditor could learn the data content by data block retrieval in the proof of possession step during the data audit procedure [67].

Several studies [68-71] introduced privacy protection protocols to determine the privacy leakage issues. In [69], the researchers designed a zero-knowledge proof-based privacy-preserving scheme with publicly provable data. Their method saves the computational and communication overhead compared to [72] and ensures data integrity without exposing the content of the user's data to the third-party auditor. However, it does not offer dynamic data operations and is limited to static data only. The above literature shows that the currently proposed data auditing schemes use the public key cryptography approach. So, the cloud needs to identify the user before hosting the data on remote storage servers for spam prevention.

Wang et al. [73] addressed this problem and suggested an ID-based outsourced data verification technique (ID-RDPC) to overcome this problem. It is the first scheme that considers being secured under Diffie-Hellman supposition. Their strategy is more optimized for communication overhead by comparing with [74]. However, their scheme cannot support public audit, generating an extra burden of computational and communication costs at the user-side during the data verification step. Thus, the ID-RDPC scheme is not suitable for resource-constrained devices.

Similarly, Zhang et al. [75] presented another user's identity-based remote data verification protocol using homomorphic tags. In their technique, the computational overhead remains consistent for the third-party auditor for the number of challenge queries. Moreover, their scheme is more proficient for computational and communicational overheads than [47, 76].

Jiang et al. [77] introduced a publically verifiable data integrity auditing technique with client revocation. They utilized a public-key cryptography database and aggregate signature to overcome collusion issues between the cloud and revoked users. It allows users to host data to the storage server and generates authentication codes to verify users from the revocation list. They also enhanced their technique with a batch auditing mechanism, which is challenging to implement in public auditing schemes. Besides, it incorporates other security features, for example, confidentiality, accountability, and traceability, to secure group user updates. The third-party auditor cannot authenticate the impersonation attack if the user does not exist in the revoked list.

Similarly, Fu et al. [78] were motivated to design a new external auditing approach to share data with group managers in CSP. They used standard MHT to maintain verifiable data blocks. This technique guarantees the users to trace the changes through an assigned hash-based binary tree and recover the most recent block of data when any existing data block corrupts. However, it is susceptible to numerous assaults, such as tag forgery, replace, replay, pollution, and data leakage attacks. Later, an identity-based cloud storage technique for the clients to remotely store their data was securely presented by Wang et al. [79]. Their approach does not need to manage certificates and enables inclusive data auditing. They also permit the proxy server to process and host the data file on the user's behalf for efficiency. However, their technique does not support data privacy and recovery, treating all system parties as trusted entities.

A scheme to secure outsourced data in the cloud uses the re-computing codes suggested by Liu et al. [80]. A semi-authorized proxy server is accessible on the user’s behalf. The proxy server keeps the compensation of the hashed data blocks and homomorphic verifiers dependent on BLS-based signatures. Furthermore, the proxy-server resolves unapproved validator issues generated by specific keys in the absence of a user who is not consistently accessible online. Consequently, this technique could nearly release the burden of data owners to become available online permanently. They used the coefficients encoding and a pseudorandom method to accomplish the privacy of the user’s data. Nonetheless, their strategy is unable to support external data auditing.

Yuan and Yu [81] presented a new public data auditing PoR mechanism based on homomorphic linear authentication commitment with constant communication costs for the cloud and TPA. It provides proficiency for storage, communication, and computational costs and releases the user's burden from being always online for data auditing. Though, it cannot support dynamic data update procedures, batch auditing, preserving data privacy, and Blockless verification.

Later, Fan et al. [82] introduced an ID-based data verification technique using aggregate signatures named (SIBAS) to overcome the vulnerability of user’s data to an adversary CSP. They introduce a trusted auditor (TEE) to verify the remotely stored data on the local side. Moreover, the scheme also accomplishes the management of security keys in the TEE environment by using Shamir’s (t, n) mechanism. It used group Diffie-Hellman supposition under the random oracle model (ROM) to resist the adversary attacks that might select messages and target identities for security verification. It is also optimized for computation and communication costs compared to [38, 76]. However, it does not support privacy-preserving, Blockless validation, and dynamic data operation.

In study [83], the researchers presented a secure data deduplication and integrity verification protocol, which can decrease the data volume hosted on cloud storage by removing identical copies of the data file. It also allows users to delegate the computational procedures to the trusted third-party auditor to authenticate data integrity proficiently. Several research works have been directed toward these issues, while this study designs a new scheme by combining the features like deduplication and publicly provable data integrity. This study supports the third-party auditor in releasing the user-side burden, particularly for resource-constrained mobile devices. Moreover, it used the linear-homomorphic authenticators and BLS signature to perform challenge-response mechanisms. However, it generates high computational and communication costs in the data deduplication step on the user side. It does not provide data update operations and batch auditing.

A multi-agent and multi-copy-based data integrity authentication technique for big data files in single cloud storage servers with low audit efficiency was introduced by Chunbo Wang and Xiaoqiang Di [84]. It utilizes a bilinear mapping technique to build a key creation procedure and multi-branch confirmation tree for performing multi-copy data signature to deploy multi-copy validation, signature, and confirmation. Moreover, it addresses task association in the work process, task assignment, and asset allotment dependent on QoS request inclination settings to plan various tasks using a directed acyclic graph (DAG). Moreover, it reduces the communication cost and storage overhead and improves audit proficiency by 20% compared to [63, 85]. However, it cannot perform dynamic data updates, batch auditing to increase efficiency, and privacy-preserving against TPA.

Yang et al. [86] introduced a certificate-less signature-based scheme for multi-user privacy-preserving with traceability and confirmation for cloud data auditing. They addressed denial-of-service (DoS) attacks, single-supervisor misconduct, and identity revelation problems. In contrast to the conventional data integrity techniques, it preserves the secrecy of the user’s identity without using a group and ring signature, which ensures the tag is minimal. Besides, it supports collaborative traceability of malevolent users with at least d managers, evading single-supervisor power maltreatment. It overcomes the DoS attack between TPA and cloud server providers by using identity verification measures. Any user can send a challenge to CSP to resolve the network congestion and waste of cloud resources. It also supports proficient user revocation and releases the burden of certificate management and key-escrow problems using certificate-less cryptography. However, TPA cannot prevent impersonation attacks if the user does not exist in the revocation list, imposing high communication and computation costs due to static data.

Later, a study [87] presented an efficient PDP scheme under the Diffie-Hellman assumption to verify data integrity in storage servers by preserving users' anonymity against TPA. Therefore, the auditor cannot get the user’s identity in the data audit process. It avoids certificate management by using an identity-based cryptographic approach. It ensures the connection between data and the owner in the proof creation step, not the integrity audit phase. Hence, TPA is unable to know liaison to find challenged data usage. Simultaneously, the CSP creates a relation for the proofs in the initialize phase to diminish TPA's computational overhead significantly. Moreover, use arbitrary requested data blocks in the proofs step to strengthen the security of the technique. Though, it is unable to support dynamic data updates and batch auditing.

Neela et al. [88] introduced a technique by improving Rivest-Shamir-Adleman (RSA) algorithm with Cuckoo Filter for secure cloud storage in semi-trusted CSPs. They eliminate the third-party auditor to overcome privacy issues and reduce communication overhead. However, they impose high computation costs on the user side and cannot support public data auditing. Later, Chaudhari et al. [89] suggested the data auditing technique based on modern Indistinguishability Obfuscation, a modern encryption construct that used one-way hash functions. The main goal of their study is to address public verification, dynamic data, collusion resistance, and privacy-preserving. Though, it cannot support batch auditing to reduce computation overhead at TPA, including high verification time, especially for mobile devices.

IV. Comparison of Data Integrity Auditing Techniques

This section offers a comprehensive comparative analysis of the techniques discussed in Section III, focusing on the strengths, weaknesses, and key attributes of remote data integrity audit protocols. The analysis is systematically presented in Table I and Table II. Table I details various data integrity schemes, examining their advantages and disadvantages. Meanwhile, Table II outlines the essential characteristics of state-of-the-art data integrity auditing protocols. These characteristics encompass Public Auditing, Dynamic Data Support, Privacy Preservation, Blockless Verification, Support for Unlimited Queries, Batch Verification, and Data Sharing capabilities.

TABLE I. Merits and Demerits of Data Integrity Schemes

Schemes	Merits	Demerits
Ateniese et al. [34]	Used RSA-based homomorphic encryption to provide public data auditing. Their S-PDP shame ensures data possession, and E-PDP improved proficiency with weak data possession.	This work does not address dynamic data operation. Hence, direct extension raises security, privacy, and system design issues.
Ateniese et al. [60]	Improve the PDP protocol proposed by [34] and provide some dynamic data operations.	It is limited to the number of challenges during the audit phase and supports only limited dynamic data operations; for instance, unable to support block insertion operations.
Bowers et al. [61]	Improves proficiency and security of the existing techniques and efficiency against the active mobile adversary.	Restricted to static data and required to store one segment of each file locally, limited to the number of challenge queries.
Zhu et al., [90]	It used the same construction of a message authentication code proposed in [61] by combining universal hashing with PRFs.	It is unappropriated for the system that requires substantial data read operations. It minimized the storage cost by increasing the communication overhead.
Xu et al. [62]	They considered the private data auditing and efficiency issues of the scheme designed by [63] and provided an efficient public data auditing protocol.	Their protocol is restricted to static data only; it does not consider the privacy leakage to the TPA
Stefanov et al. [64]	Iris supports publicly verifiable dynamic data auditing by using the MHT. It provides a proficient read and writes data operation.	It requires substantial bandwidth, server computation time, local data space, and server input/output to verify the data integrity.
Shi et al. [65]	It supports efficient reading and writing operations and provides a publicly provable dynamic nature of data using standard MHT.	Validate the integrity requires the high computational power of the cloud server, bandwidth, and local storage space.
Zhang et al. [70]	It provides privacy-preserving public data auditing PoR-based scheme with aggregate verification.	It is not suitable for dynamic data environments and is only limited to static files.
Yu et al. [69]	It provides a zero-knowledge proof-based method to prevent data privacy leakage in the audit process.	The proposed scheme is limited to static data only.
Tan and Jia [76]	Identity-based cumulative signature is used to generate homomorphic tags. Eliminates data auditing burden on cloud users and is proficient to computational and communicational overheads compared with [47, 76].	It cannot perform dynamic data operations; the TPA can expose data content during the audit phase, which raises privacy issues.
Wang et al. [73]	It generates homomorphic aggregate tags using the user identity; it is optimized in terms of communication and computation cost compared with [34, 74].	Not suitable for resource constraints devices because it supports private audit, which generates computation and communication overheads on user-side in the audit phase.
Jiang et al. [77]	Used public-key cryptography and group signature to overcome collusion issues between the cloud and revoked users, generating the authentication code to verify the user from the revocation list.	It is unable to provide data recovery, imposes high computation and communication overhead, and the auditor is not capable of verifying impersonation attacks.
Yuan and Yu [81]	Used homomorphic authenticators with constant communication for cloud and TPA. Proficient in computation cost and reduced the storage overhead.	It is unable to provide dynamic data procedures.
Liu et al. [80]	Proxy resolves the user’s absence issue by generating specific keys, using the coefficients encoding and a pseudorandom method to accomplish the privacy of the user’s data.	Their approach is unable to support public data audits.
Fu et al. [78]	Use MHT to preserve verifiable data blocks. Ensures users trace the changes and recover the most recent block of data when an existing data block corrupts.	It is vulnerable to tag forgery, replace, replay, pollution, data leakage attacks, etc.
Wang et al. [79]	It does not need to manage certificates and enables inclusive data auditing, more proficient by incorporating a proxy server for processing and hosting the data file on the user’s behalf.	It does not provide data privacy or recovery, and all the involved entities in the system are treated as trusted.
Fan et al. [82]	Using aggregate signatures to overcome vulnerability against untrusted CSP resists adversary attacks that select its data and target identities, optimizing computation and communication costs more than [38, 76].	Restricted to the static nature of data, privacy-preserving, Blockless verification, dynamic data operation, replay, replace, tag-forgery attacks, etc.
Youn et al. [83]	It performs secure deduplication and data integrity verification, reduces the storage overhead by removing duplicated copies, and uses linear-homomorphic authenticators with BLS signature to perform challenge-response mechanisms.	Generating high communication and computation costs on the user-side during the data deduplication phase is unsuitable for mobile device resource constraints. It does not support dynamic data updates and batch auditing.
Wang and Di [84]	it reduces the communication cost and storage overhead and improves audit proficiency by 20% compared to [63, 85].	It is unable to perform dynamic data updates, batch auditing to increase efficiency, and privacy-preserving against TPA.

Yang et al. [86]	Provides collaborative traceability of malevolent users to minimize the network congestion and waste of cloud resources, preserves identity revelation and DoS attack, and uses certificate-less cryptography.	The auditor cannot authenticate impersonation attacks if the user does not exist in the revoke list, imposing high communication and computation costs due to the static data.
Yan and Gui [87]	The auditor is unable to get the user's identity in the data audit process; it avoids the certificate management and cloud setting up a connection in the proofs creation phase to minimize the computation cost of TPA.	It is unable to provide data-dynamic and batch auditing.
Neela et al. [88]	Improve RSA algorithm with Cuckoo Filter for secure storage in semi-trusted CSPs.	Impose high computation costs on the user side and cannot support public data auditing.
Chaudhari et al. [89]	The auditor cannot get the user's identity in the data audit process; it avoids the certificate management and cloud setting up a connection in the proofs creation phase to minimize the computation cost of TPA.	It does not support batch auditing and imposes high computation and communication costs, which is unsuitable for energy-constrained mobile devices.

TABLE II. DESIRABLE FEATURES FOR STATE-OF-THE-ART DATA INTEGRITY AUDITING PROTOCOLS

Schemes	**Auditing Requirements**						
	Public Auditing	Dynamic Data	Privacy-Preserving	Blockless Verification	Unlimited Queries	Batch Verification	Data Sharing
Ateniese et al. [34]	✓	✗	✗	✓	✗	✗	✗
Ateniese et al. [60]	✗	✓	✓	✗	✗	✗	✗
Bowers et al. [61]	✗	✗	✗	✗	✗	✗	✗
Zhu et al. [90]	✗	✓	✗	✓	✓	✗	✗
Xu et al. [62]	✓	✗	✗	✗	✗	✗	✗
Stefanov et al. [64]	✗	✓	✗	✗	✗	✗	✗
Shi et al. [65]	✗	✗	✗	✗	✗	✗	✗
Zhang et al. [70]	✓	✗	✓	✗	✗	✗	✗
Yu et al. [69]	✗	✗	✓	✗	✗	✗	✗
Tan and Jia [76]	✓	✗	✗	✓	✓	✗	✗
Wang et al. [73]	✗	✗	✗	✗	✗	✗	✗
Jiang et al. [77]	✓	✗	✗	✗	✗	✓	✓
Yuan and Yu [81]	✓	✗	✗	✗	✓	✗	✓
Liu et al. [80]	✓	✗	✗	✓	✓	✗	✗
Fu et al. [78]	✓	✗	✗	✗	✗	✗	✓
Wang et al. [79]	✓	✗	✗	✗	✗	✗	✗
Fan et al. [82]	✓	✗	✗	✗	✗	✓	✗
Youn et al. [83]	✓	✗	✗	✗	✓	✗	✗
Wang and Di [84]	✓	✗	✗	✓	✓	✗	✗
Yang et al. [86]	✓	✗	✗	✗	✓	✗	✓
Yan and Gui [87]	✓	✗	✗	✓	✓	✗	✓
Neela et al. [88]	✗	✓	✗	✗	✓	✗	✗
Chaudhari et al. [89]	✓	✓	✓	✗	✗	✗	✗

V. PERFORMANCE-BASED COMPARISON TECHNIQUES

A performance comparison to evaluate different protocols and data structure involved in these schemes are presented in this section. The following section will explain several open research issues to design a proficient data auditing protocol.

The cloud data storage is not indisputable and is constantly modified according to the user's request, for example, insertion, deletion, modification, and append procedures. For example, when the user's first-time outsourced data may not be complete, the user needs to update and complete it after uploading it to CSP. Alternatively, the client may want to delete obsolete or useless data from cloud storage servers, which is unavoidable for utilizing cloud services. In short, data update operations are necessary to protect cloud data storage, and several techniques have been developed for this perspective. Generally, the subsequent data techniques are commonly utilized to develop data update verification techniques for outsourcing users' data.

A. *Merkle Hash Tree (MHT)*

It is a famous binary hash tree data structure introduced by [42]. It can proficiently ensure and verify whether a series of data blocks is compromised or not. The tree nodes hold the hash value of relating sibling data blocks. The MHT is designed by computing the has value in pair from bottom to up and obtaining the root node's distinct hash value. The MHT is a provable and broadly explored data structure supporting data dynamics operations. The MHT-based secure cloud data verification technique was suggested in [42].

Nevertheless, the curious storage servers may pass the audit process without appropriate authentication of the block indices by generating proof with other authentic data blocks when the assaulted data blocks are compromised. In [50], researchers proposed a fine-grained data dynamics technique for remotely storing user's data. However, they assume that the cloud storage server remains honest in responding to the challenging queries over outsourced data. Later, in [53], researchers utilized the MHT in data duplication auditing; it develops an MHT for every copy of the data file and uses the root node to construct the two-level data structure. However, in this way, the usage of MHT cannot support sequence number authentication issues.

To overcome these issues, the study [51] proposed a multi-replica technique that integrates the other parameters, including node level and node number accessible from the node. Boolean value denotes the node either on left or right node location to its parent node in auxiliary authentication path. In this manner, multiple-replica MHT can design to validate data update operations and authenticate the block indices proficiently.

B. Ranked-Based Skip List (RASL)

It is an authentication model that verifies the integrity of data blocks and provides data dynamics operations. It is a hierarchical key-value pair storage data structure like a tree, where nodes are arranged concerning their corresponding keys. By using this technique, every node v stores two parameters $right(v)$ and $down(v)$ with the goal that a particular data node can position in searching procedure. The study [91] introduced the RSAL-based first PDP technique. In their proposed structure, every node stores $right(v)$, $down(v)$, and produced $f(v)$ by iteratively using hash method to $f\left(right(v)\right)$ and $f\left(down(v)\right)$. The root value use to validate server's response in challenge-proof phase. The primary disadvantage of this technique is the absence of checking the single block integrity. In [41], the researchers enhanced the RSAL to multi-replica data dynamics. The enhancement integrates all data blocks of all data copies in an identical structure, supporting the proficient data auditing mechanism for copies. Moreover, using M-RASL can minimize the communication cost for updating authentication of outsourced data with multi-replicas.

C. Index Hash Table (IHT)

IHT stores the data block modification and supports to creation hash value of every data block during authentication. The IHT design is identical to an array, which involves indices sequence i_{no}, block sequence b_{no}, block version sequence v_{no} and arbitrary value r_{no}. It was first proposed in [85] for cloud storage, decreasing computational and communicational costs by keeping the index hash table on the auditor rather than the storage server. Insertion and deletion procedures reason for the change of b_{no}, thus suffering re-generation of compromised data blocks.

A subsequent procedure is to change the arrangement of specific components, such as $(i_{no}, b_{no}, v_{no}, r_{no})$ [31]. In [47], researchers modified the components to (i_{no}, V_i, R_i, S_i), where V_i denotes the virtual index, S_i represents the signer identity of the data block. In particular, guarantees that every data block is in the proper order, and V_i of the inserted data block m_i' use the smallest number of $m_i' = (V_i + V_{i-1})/2$. Thus, the user can adequately run the data insert procedure for shared data.

Additionally, there are two more distinctive data structures, like Dynamic Hash Table (DHT) [92] and Novel Data Structure (NDS), introduced by [49], respectively. Endeavoring enhanced data update operations proficiency, a DHT-based publicly provable data auditing technique was introduced to ensure files and block-level modification. The DHT is innovative two-dimensional information stored by the auditor, minimizing communicational costs during data auditing and modification authentication.

Even though it provides better results than the existing auditing techniques, it still has a few deficiencies. First, the cloud storage server may suffer a collision attack between user and auditor because the user creates a time-stamp for validation. The auditor has just approved the user. Furthermore, it does not integrate the indices and position sequence of the data blocks. To address these problems, NDS was intended to support data dynamics operations comprising doubly-link information and position array.

The user with a novel data structure can perform insertion and deletion procedures without affecting other data blocks. Likewise, NDS effectively controls the connection between the given blocks of data and their particular position can also be appropriate for multi-copies. The user may change auditor inappropriately due to geographic location or price effect. NDS should be re-structured for every change, leading to extra overhead compared to implementing the developed data structure stored by the cloud server.

VI. Open Research Issues and Challenges

Several RDA schemes are currently proposed to ensure outsourced users' data in cloud environments. However, some problems that require attention as an open research direction are described below:

A. Dynamic Data Update

In the static data auditing approach, to change the location or update a single data bit, the user needs to modify more than half bits of files after downloading the entire data and again hosted on the cloud storage server. It creates high computation, communication, and storage overheads for the user, cloud, and the auditor. Consequently, dynamic data operation is a fundamental property of the outsourced data integrity audit techniques, for example, electronic records and log files in cloud computing and mobile cloud computing. Nonetheless, the main restriction of PoR and POW-based schemes is the data updates procedure [91].

Even though the study [93] introduced a PoR-based scheme for avoiding data update issues in cloud infrastructure, private data auditing makes this technique unusable for mobile cloud computing. Furthermore, existing dynamic data update strategies also suffer computation costs on the user side, particularly for resource constraints devices. Accordingly, allowing mobile clients to update their data effectively requires future research and improvements.

B. Shared Data Access Control

Currently, several CSPs offer services, including online blogs, web services, and web applications required to store their data on remote cloud storage servers. These clients must gain access to their data anytime, anywhere, and simultaneously execute data update operations. For example, most hosted websites support their users to update data freely. Nonetheless, most existing techniques guaranteeing data integrity cannot fully achieve these requirements or impose high computation and storage overhead on the user side.

C. Data Privacy Issue

Supporting data privacy has been essential to meeting the SLA for cloud storage services. A public data auditing approach must not expose data privacy to third-party auditors. An auditor must be proficient in securely conducting the audit process regardless of any security risk of exposing the data content. For privacy-preserving, a MAC-based solution can use for users' data. The TPA generates cloud storage challenges for data integrity by arbitrarily selecting data blocks and respective MAC. The cloud storage replies with data blocks and MAC, and then the TPA verifies it using the secret key. However, as the server responds, a linear data blocks sequence to TPA, so the auditor can breach the SLA between the user and TPA. Moreover, it supports static data and restricts the number of challenges. The user must retrieve the complete data, generate the new keys, and host it on the storage servers, imposing significant computation and communication costs. Homomorphic authenticators with arbitrary masking can use to overcome these issues [38].

D. Blockless Data Auditing

An audit strategy that does not utilize accumulation signature and verification tag need the cloud server to respond to the requested data blocks for integrity assurance. This approach imposes a high communication cost on the cloud server and affects audit proficiency. The study Wang et al. [42] introduced a Blockless auditing approach that verifies tags rather than validating actual data blocks in the auditing phase. Even though this approach can improve the proficiency of the data auditing strategy and highly minimizes the communication cost, it might allow the CSP to cheat. Assume the user wants to achieve the data update procedure, which is possible because the storage server possesses old data and signatures. As the signatures and data are legitimate, the auditor might not recognize whether they are appropriately updated or not.

E. Deletion Assurance

It is required to assure the data delete operation; else, it might result in a data breach for the cloud storage [94]. Another significant perspective in guaranteed data deletion is that different versions of data files with similar data contents of the deleted identical copies or files must be kept secure. Rahumed et al. [95] introduced a fine-grained version control backup system named "FADE" to ensure deletion operation. The deleted versions of files are guaranteed to be forever unavailable. In [96], the researchers describe mechanisms for securely assuring data deletion, the classification of adversaries, and capacities for exploiting the cloud servers without secured deletion. This term can also be named self-destruction in [97],[98],[99] and [100].

VII. Future Research Work

Cloud computing has a rapidly growing business model and evolving new features to facilitate users. The researchers can take advantage to improve the data auditing schemes. The future research direction concerning data auditing with cloud computing is following.

A. Geolocation Assurance

The CSPs are restricted by Service Level Arrangement (SLA) to outsource users' data in a specific geographic area with a particular time zone, political border, state, or city level. For administration reasons and law variance, some cloud clients needed to guarantee data geolocation [101], [102], [103] and [104]. The CSP may transfer users' data to abroad servers for less expensive IT costs, disregarding the SLA agreement. Such activities of the CSP may disclose the client's data to foreign governments, who can assess it via court orders or any lawful approach. In this situation, timely recognition proof of purposeful deceitfulness or breach of SLA with CSPs is a fundamental need for cloud clients. The data integrity verification technique must incorporate geolocation confirmation for future research work.

B. Big Data Auditing

Big data auditing of RDA schemes is challenging because storage, communication, and computation costs may exponentially increase when data size increases. The cloud providers may also delete rarely used data to save storage costs without intimating the user and try to hide data loss info for the sack of their repute [105], [106], [107] and [108]. Consequently, it needs to develop an RDA scheme for supporting big-data auditing to minimize storage overhead, computation complexity, and communication cost.

C. Support for Collaborative Auditing

To audit the user's data in a multi-cloud environment is named collaborative auditing. Several data auditing schemes [34], [38], [42], [44], [60], [63], and [91] proposed for a single cloud cannot support a multi-cloud environment. Distributed File Systems were introduced to cloud storage systems for local independence and low cost for owner's data. However, collaborative auditing is challenging for job assignments, security assurance, and communication costs. Furthermore, reducing the computation overhead, storage cost, and system usability should also be considered when designing data auditing protocols.

D. Data Auditing with Deduplication

The cloud server store duplicates data, especially in multi-replica that generates different copies of the same data file, creating storage overhead for identical copies. It is significant to resolve this problem by improving the RDA schemes. Currently, [109], [110], and [111] were suggested data auditing techniques with PoW to overcome this issue considering only single data files. However, how to improve the multi-replica RDA scheme with PoW mechanism to ensure data security still seems a vital research contribution in the future.

E. Blockchain-Based Data Auditng

The blockchain is an undisputable distributed accounting, consensus approach, and intelligent contract technology with asymmetric cryptography, ensuring data security and privacy. The researchers can take advantage of decentralizing features that could generally use in data integrity auditing schemes [112], [113] and [114] for single and distributed infrastructures. However, improving the security framework of data auditing techniques for single and multiple copies scenario could be an auspicious future research direction.

F. Data Auditing in Fog Computing

Fog computing is an extension of cloud computing to facilitate the edge network for easy access and fast corresponding services for end-users. Users can use multiple devices as fog nodes to support cloud servers and reduce CSP charges. Recently, Wang et al. [115] suggested a fog-based secure and anonymous data auditing scheme. Even though the scheme is efficient and secure; however, they consider the cloud a fully trusted entity. Therefore, secure fog-based data auditing is still an open research direction for cloud computing.

G. Data Auditing in Edge Computing

Edge computing is a distributed computing paradigm that provides computational and storage facilities closer to the end-user for fast response times (low latency) and reducing bandwidth utilization. It is an ongoing business model that requires improving data auditing techniques to incorporate cloud computing changes. However, duplicated data may lead to storage problems for single and shared clouds; auditing techniques must include deduplication [116] and [117] and ensure data security and privacy in edge computing. Furthermore, the researchers can explore the edge computing model for data auditing with blockchain technology in future research. Cloud computing saves time and monitoring costs for any organization and turns technological solutions for large-scale systems into server-to-service frameworks [118].

VIII. Conclusion

The paper has highlighted the significance of remote data auditing techniques, involved entities, and the concerns related to public and private data auditing. We thoroughly reviewed existing data integrity techniques by exploring their merits and demerits. Also, we introduced the qualitative comparison of auditing techniques and compared their performance regarding communication and computational overhead. Lastly, we highlighted the desirable features of different schemes, open issues, and future research directions for designing efficient and secure data auditing schemes for the cloud environment.

Acknowledgment

This research was supported by Ministry of Higher Education (MoHE) of Malaysia through Fundamental Research Grant Scheme (FRGS/1/2018/ICT04/UUM/02/17).

References

[1] X. Ma, H. Gao, H. Xu, and M. Bian, "An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing," EURASIP Journal on Wireless Communications Networking, vol. 2019, no. 1, pp. 1-19, 2019.

[2] Y. Yin, L. Chen, Y. Xu, J. Wan, H. Zhang, and Z. Mai, "QoS prediction for service recommendation with deep feature learning in edge computing environment," Mobile Networks and Applications, pp. 1-11, 2019.

[3] H. Gao, L. Kuang, Y. Yin, B. Guo, and K. Dou, "Mining consuming behaviors with temporal evolution for personalized recommendation in mobile marketing apps," Mobile Networks and Applications, vol. 25, pp. 1233-1248, 2020.

[4] Linode. "Linode LCC company." https://www.linode.com/ (accessed May 25, 2021).

[5] Amazon. "Amazon Elastic Compute Cloud." https://aws.amazon.com/ec2/ (accessed 24 May, 2021).

[6] Google. "Accelerate your transformation with Google Cloud." https://cloud.google.com/ (accessed 24 May, 2022).

[7] Azure. "Microsoft Azure." https://azure.microsoft.com/en-us/ (accessed March 14, 2022).

[8] X. Yang, S. Zhou, and M. Cao, "An approach to alleviate the sparsity problem of hybrid collaborative filtering based recommendations: the product-attribute perspective from user reviews," Mobile Networks and Applications, pp. 1-15, 2019.

[9] H. Gao, C. Liu, Y. Li, and X. Yang, "V2VR: reliable hybrid-network-oriented V2V data transmission and routing considering RSUs and connectivity probability," IEEE Transactions on Intelligent Transportation Systems, 2020.

[10] P. Mell and T. Grance, "The NIST definition of cloud computing," 2011.

[11] E. Simmon, "Evaluation of cloud computing services based on NIST SP 800-145," NIST Special Publication, vol. 500, p. 322, 2018.

[12] C. M. Mohammed and S. R. Zebaree, "Sufficient comparison among cloud computing services: IaaS, PaaS, and SaaS: A review," International Journal of Science Business, vol. 5, no. 2, pp. 17-30, 2021.

[13] S. Srinivasan, Cloud computing basics. Springer, 2014.

[14] A. Alrabea, "A modified Boneh-Lynn-Shacham signing dynamic auditing in cloud computing," Journal of King Saud University-Computer Information Sciences, 2020.

[15] I. E. Baciu, "Advantages and disadvantages of cloud computing services, from the employee's point of view," National Strategies Observer No.2/Vol.1, 2015, vol. 2, 2015.

[16] Statista, "Public cloud services end-user spending worldwide from 2017 to 2022," 2019. [Online]. Available: https://www.statista.com/ statistics/ 273818/global-revenue-generated-with-cloud-computing-since-2009/.

[17] B. Nedelcu, M.-E. Stefanet, I.-F. Tamasescu, S.-E. Tintoiu, and A. Vezeanu, "Cloud computing and its challenges and benefits in the bank system," Database Systems Journal, vol. 6, no. 1, pp. 44-58, 2015.

[18] B. Wang, B. Li, and H. Li, "Oruta: Privacy-preserving public auditing for shared data in the cloud," IEEE transactions on cloud computing, vol. 2, no. 1, pp. 43-56, 2014.

[19] K. Thomas. "Microsoft cloud data breach heralds things to come." https://www.pcworld.com/article/214775/microsoft_cloud_data_breach_sign_of_future.html (accessed July 2021).

[20] I. Orton, A. Alva, and B. Endicott-Popovsky, "Legal process and requirements for cloud forensic investigations," in Cybercrime and Cloud Forensics: Applications for Investigation Processes: IGI Global, 2013, pp. 186-229.

[21] R. Yeluri and E. Castro-Leon, "Cloud computing basics," in Building the infrastructure for cloud security: Springer, 2014, pp. 1-17.

[22] R. Ko, S. G. Lee, and V. Rajan, "Cloud computing vulnerability incidents: A statistical overview," Cloud Security Alliance, 2013.

[23] ISACA, "Isaca Glossary," 2015.

[24] A. Team. "Amazon S3 availability event." https://status.aws.amazon.com/s3-20080720.html (accessed April 14, 2022).

[25] A. iCloud. "Apple iCloud." https://www.icloud.com/ (accessed May 29, 2021).

[26] T. Mather, S. Kumaraswamy, and S. Latif, Cloud security and privacy: an enterprise perspective on risks and compliance. " O'Reilly Media, Inc.", 2009.

[27] S. Yu, W. Lou, and K. Ren, "Data security in cloud computing," Morgan Kaufmann/Elsevier, Book section, vol. 15, pp. 389-410, 2012.

[28] M. Bellare, R. Canetti, and H. Krawczyk, "Keying hash functions for message authentication," in Annual international cryptology conference, 1996: Springer, pp. 1-15.

[29] M. Zhou, R. Zhang, W. Xie, W. Qian, and A. Zhou, "Security and privacy in cloud computing: A survey," presented at the 2010 Sixth International Conference on Semantics, Knowledge and Grids, 2010.

[30] G. Ateniese, S. Kamara, and J. Katz, "Proofs of storage from homomorphic identification protocols," presented at the International conference on the theory and application of cryptology and information security, 2009.

[31] K. Yang and X. Jia, "An efficient and secure dynamic auditing protocol for data storage in cloud computing," IEEE transactions on parallel and distributed systems, vol. 24, no. 9, pp. 1717-1726, 2013.

[32] C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for secure cloud storage," IEEE transactions on computers, vol. 62, no. 2, pp. 362-375, 2013.

[33] C. Gentry, "Fully homomorphic encryption using ideal lattices," presented at the Proceedings of the forty-first annual ACM symposium on Theory of computing, 2009.

[34] G. Ateniese et al., "Provable data possession at untrusted stores," presented at the Proceedings of the 14th ACM conference on Computer and communications security, 2007.

[35] L. A. B. Silva, C. Costa, and J. L. Oliveira, "A common API for delivering services over multi-vendor cloud resources," Journal of Systems and Software, vol. 86, no. 9, pp. 2309-2317, 2013.

[36] D. Boneh and D. M. Freeman, "Homomorphic signatures for polynomial functions," presented at the annual international conference on the theory and applications of cryptographic techniques, 2011.

[37] A. L. Ferrara, M. Green, S. Hohenberger, and M. Ø. Pedersen, "Practical short signature batch verification," in Cryptographers’ Track at the RSA Conference, 2009: Springer, pp. 309-324.

[38] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," presented at the Infocom, 2010 proceedings ieee, 2010.

[39] R. Curtmola, O. Khan, R. Burns, and G. Ateniese, "MR-PDP: Multiple-replica provable data possession," presented at the the 28th international conference on distributed computing systems, 2008.

[40] Y. Zhang, J. Ni, X. Tao, Y. Wang, and Y. Yu, "Provable multiple replication data possession with full dynamics for secure cloud storage," Concurrency Computation: Practice Experience, vol. 28, no. 4, pp. 1161-1173, 2016.

[41] A. Abo-alian, N. L. Badr, and M. F. Tolba, "Integrity as a service for replicated data on the cloud," Concurrency Computation: Practice Experience, vol. 29, no. 4, p. e3883, 2017.

[42] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," IEEE transactions on parallel and distributed systems, vol. 22, no. 5, pp. 847-859, 2011.

[43] D. Boneh, C. Gentry, B. Lynn, and H. Shacham, "A survey of two signature aggregation techniques," ed: Citeseer, 2003.

[44] A. Juels and B. S. Kaliski Jr, "PORs: Proofs of retrievability for large files," presented at the Proceedings of the 14th ACM conference on Computer and communications security, 2007.

[45] H. Wang, "Proxy provable data possession in public clouds," IEEE Transactions on Services Computing, vol. 6, no. 4, pp. 551-559, 2013.

[46] H. Wang, Q. Wu, B. Qin, and J. Domingo-Ferrer, "FRR: Fair remote retrieval of outsourced private medical records in electronic health networks," Journal of biomedical informatics, vol. 50, pp. 226-233, 2014.

[47] B. Wang, B. Li, and H. Li, "Panda: Public auditing for shared data with efficient user revocation in the cloud," IEEE Transactions on services computing, vol. 8, no. 1, pp. 92-106, 2015.

[48] G. Wu, Y. Mu, W. Susilo, and F. Guo, "Privacy-preserving cloud auditing with multiple uploaders," presented at the International Conference on Information Security Practice and Experience, 2016.

[49] J. Shen, J. Shen, X. Chen, X. Huang, and W. Susilo, "An efficient public auditing protocol with novel dynamic structure for cloud data," IEEE Transactions on Information Forensics and Security, vol. 12, no. 10, pp. 2402-2415, 2017.

[50] C. Liu et al., "Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 9, pp. 2234-2244, 2014.

[51] C. Liu, R. Ranjan, C. Yang, X. Zhang, L. Wang, and J. Chen, "MuR-DPA: Top-down levelled multi-replica merkle hash tree based secure public auditing for dynamic big data storage on cloud," IEEE Transactions on Computers, vol. 64, no. 9, pp. 2609-2622, 2015.

[52] A. F. Barsoum and M. A. Hasan, "Provable possession and replication of data over cloud servers," Centre For Applied Cryptographic Research , University of Waterloo, Report, vol. 32, p. 2010, 2010.

[53] A. F. Barsoum and M. A. Hasan, "On Verifying Dynamic Multiple Data Copies over Cloud Servers," IACR Cryptol. ePrint Arch., vol. 2011, p. 447, 2011.

[54] H. Wang, Q. Wu, B. Qin, and J. Domingo-Ferrer, "Identity-based remote data possession checking in public clouds," IET Information Security, vol. 8, no. 2, pp. 114-121, 2014.

[55] H. Wang, "Identity-based distributed provable data possession in multicloud storage," IEEE Transactions on Services Computing, vol. 8, no. 2, pp. 328-340, 2015.

[56] Y. Yu et al., "Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage," IEEE Transactions on Information Forensics Security, vol. 12, no. 4, pp. 767-778, 2016.

[57] B. Wang, B. Li, H. Li, and F. Li, "Certificateless public auditing for data integrity in the cloud," presented at the 2013 IEEE conference on communications and network security (CNS), 2013.

[58] D. He, S. Zeadally, and L. Wu, "Certificateless public auditing scheme for cloud-assisted wireless body area networks," IEEE Systems Journal, vol. 12, no. 1, pp. 64-73, 2015.

[59] D. He, N. Kumar, H. Wang, L. Wang, and K.-K. R. Choo, "Privacy-preserving certificateless provable data possession scheme for big data storage on cloud," Applied Mathematics Computation, vol. 314, pp. 31-43, 2017.

[60] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," presented at the Proceedings of the 4th international conference on Security and privacy in communication netowrks, 2008.

[61] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A high-availability and integrity layer for cloud storage," presented at the Proceedings of the 16th ACM conference on Computer and communications security, 2009.

[62] J. Xu and E.-C. Chang, "Towards efficient proofs of retrievability," presented at the Proceedings of the 7th ACM symposium on information, computer and communications security, 2012.

[63] H. Shacham and B. Waters, "Compact proofs of retrievability," presented at the International Conference on the Theory and Application of Cryptology and Information Security, 2008.

[64] E. Stefanov, M. van Dijk, A. Juels, and A. Oprea, "Iris: A scalable cloud file system with efficient integrity checks," presented at the Proceedings of the 28th Annual Computer Security Applications Conference, 2012.

[65] E. Shi, E. Stefanov, and C. Papamanthou, "Practical dynamic proofs of retrievability," presented at the Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, 2013.

[66] R. C. Merkle, "A certified digital signature," presented at the Conference on the Theory and Application of Cryptology, 1989.

[67] D. Zissis and D. Lekkas, "Addressing cloud computing security issues," Future Generation computer systems, vol. 28, no. 3, pp. 583-592, 2012.

[68] C. Li, Y. Chen, P. Tan, and G. Yang, "Towards comprehensive provable data possession in cloud computing," Wuhan University Journal of Natural Sciences, vol. 18, no. 3, pp. 265-271, 2013.

[69] Y. Yu et al., "Enhanced privacy of a remote data integrity-checking protocol for secure cloud storage," International Journal of Information Security, vol. 14, no. 4, pp. 307-318, 2015.

[70] J. Zhang, W. Tang, and J. Mao, "Efficient public verification proof of retrievability scheme in cloud," Cluster computing, vol. 17, no. 4, pp. 1401-1411, 2014.

[71] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, and H. Hu, "Zero-knowledge proofs of retrievability," Science China Information Sciences, vol. 54, no. 8, p. 1608, 2011.

[72] Z. Hao, S. Zhong, and N. Yu, "A privacy-preserving remote data integrity checking protocol with data dynamics and public verifiability," IEEE transactions on Knowledge and Data Engineering, vol. 23, no. 9, pp. 1432-1437, 2011.

[73] H. Wang, Q. Wu, B. Qin, and J. Domingo-Ferrer, "Identity-based remote data possession checking in public clouds," IET Information Security, vol. 8, no. 2, pp. 114-121, 2014.

[74] F. Sebé, J. Domingo-Ferrer, A. Martinez-Balleste, Y. Deswarte, and J.-J. Quisquater, "Efficient remote data possession checking in critical information infrastructures," IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 8, pp. 1034-1038, 2008.

[75] J. Zhang, P. Li, and J. Mao, "IPad: ID-based public auditing for the outsourced data in the standard model," Cluster Computing, vol. 19, no. 1, pp. 127-138, 2016.

[76] S. Tan and Y. Jia, "NaEPASC: a novel and efficient public auditing scheme for cloud data," Journal of Zhejiang University SCIENCE C, vol. 15, no. 9, pp. 794-804, 2015.

[77] T. Jiang, X. Chen, and J. Ma, "Public integrity auditing for shared dynamic cloud data with group user revocation," IEEE Transactions on Computers, vol. 65, no. 8, pp. 2363-2373, 2016.

[78] A. Fu, S. Yu, Y. Zhang, H. Wang, and C. Huang, "NPP: a new privacy-aware public auditing scheme for cloud data sharing with group users," IEEE Transactions on Big Data, 2017.

[79] Y. Wang, Q. Wu, B. Qin, W. Shi, R. H. Deng, and J. Hu, "Identity-based data outsourcing with comprehensive auditing in clouds," IEEE Transactions on Information Forensics and Security, vol. 12, no. 4, pp. 940-952, 2017.

[80] J. Liu, K. Huang, H. Rong, H. Wang, and M. Xian, "Privacy-preserving public auditing for regenerating-code-based cloud storage," IEEE Transactions on Information Forensics and Security, vol. 10, no. 7, pp. 1513-1528, 2015.

[81] J. Yuan and S. Yu, "Public integrity auditing for dynamic data sharing with multiuser modification," IEEE Transactions on Information Forensics and Security, vol. 10, no. 8, pp. 1717-1726, 2015.

[82] Y. Fan, X. Lin, G. Tan, Y. Zhang, W. Dong, and J. Lei, "One secure data integrity verification scheme for cloud storage," Future Generation Computer Systems, vol. 96, pp. 376-385, 2019.

[83] T.-Y. Youn, K.-Y. Chang, K.-H. Rhee, and S. U. Shin, "Efficient client-side deduplication of encrypted data with public auditing in cloud storage," IEEE Access, vol. 6, pp. 26578-26587, 2018.

[84] C. Wang and X. Di, "Research on Integrity Check Method of Cloud Storage Multi-Copy Data Based on Multi-Agent," IEEE Access, vol. 8, pp. 17170-17178, 2020.

[85] Y. Zhu, G.-J. Ahn, H. Hu, S. S. Yau, H. G. An, and C.-J. Hu, "Dynamic audit services for outsourced storages in clouds," IEEE Transactions on Services Computing, vol. 6, no. 2, pp. 227-238, 2013.

[86] X. Yang, M. Wang, T. Li, R. Liu, and C. Wang, "Privacy-Preserving Cloud Auditing for Multiple Users Scheme With Authorization and Traceability," IEEE Access, vol. 8, pp. 130866-130877, 2020.

[87] H. Yan and W. Gui, "Efficient Identity-Based Public Integrity Auditing of Shared Data in Cloud Storage With User Privacy Preserving," IEEE Access, vol. 9, pp. 45822-45831, 2021.

[88] K. Neela and V. Kavitha, "An Improved RSA Technique with Efficient Data Integrity Verification for Outsourcing Database in Cloud," Wireless Personal Communications, pp. 1-18, 2022.

[89] S. Chaudhari and G. Swain, "Towards Lightweight Provable Data Possession for Cloud Storage Using Indistinguishability Obfuscation," IEEE Access, vol. 10, pp. 31607-31625, 2022.

[90] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S. S. Yau, "Efficient provable data possession for hybrid clouds," presented at the Proceedings of the 17th ACM conference on Computer and communications security, 2010.

[91] C. Erway, A. Küpçü, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," presented at the Proceedings of the 16th ACM conference on Computer and communications security, Chicago, Illinois, USA, 2009.

[92] H. Tian et al., "Dynamic-hash-table based public auditing for secure cloud storage," IEEE Transactions on Services Computing, 2015.

[93] D. Cash, A. Küpçü, and D. Wichs, "Dynamic proofs of retrievability via oblivious RAM," Journal of Cryptology, vol. 30, no. 1, pp. 22-57, 2013.

[94] Y. Tang, P. P. Lee, J. C. Lui, and R. Perlman, "Secure overlay cloud storage with access control and assured deletion," IEEE Transactions on dependable secure computing, vol. 9, no. 6, pp. 903-916, 2012.

[95] A. Rahumed, H. C. Chen, Y. Tang, P. P. Lee, and J. C. Lui, "A secure cloud backup system with assured deletion and version control," presented at the 2011 40th International Conference on Parallel Processing Workshops, 2011.

[96] J. Reardon, D. Basin, and S. Capkun, "Sok: Secure data deletion," presented at the 2013 IEEE symposium on security and privacy, 2013.

[97] C. Cachin, K. Haralambiev, H.-C. Hsiao, and A. Sorniotti, "Policy-based secure deletion," presented at the Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, 2013.

[98] C. Li, Y. Chen, and Y. Zhou, "A data assured deletion scheme in cloud storage," China Communications, vol. 11, no. 4, pp. 98-110, 2014.

[99] A. B. Habib, T. Khanam, and R. Palit, "Simplified file assured deletion (sfade)-a user friendly overlay approach for data security in cloud storage system," presented at the 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013.

[100]J. Xiong, Z. Yao, J. Ma, X. Liu, and Q. Li, "A secure document self-destruction scheme: an ABE approach," presented at the 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing, 2013.

[101]A. Albeshri, C. Boyd, and J. G. Nieto, "Enhanced geoproof: improved geographic assurance for data in the cloud," International Journal of Information Security, vol. 13, no. 2, pp. 191-198, 2014.

[102]D. L. Fu, X. G. Peng, and Y. L. Yang, "Trusted validation for geolocation of cloud data," The Computer Journal, vol. 58, no. 10, pp. 2595-2607, 2015.

[103]M. Gondree and Z. N. Peterson, "Geolocation of data in the cloud," presented at the Proceedings of the third ACM conference on Data and application security and privacy, 2013.

[104]T. Jiang, W. Meng, X. Yuan, L. Wang, J. Ge, and J. Ma, "ReliableBox: Secure and Verifiable Cloud Storage With Location-Aware Backup," IEEE Transactions on Parallel and Distributed Systems, vol. 32, no. 12, pp. 2996-3010, 2021.

[105]L. Liu, O. De Vel, Q.-L. Han, J. Zhang, and Y. Xiang, "Detecting and preventing cyber insider threats: A survey," IEEE Communications Surveys and Tutorials, vol. 20, no. 2, pp. 1397-1417, 2018.

[106]H. Tabrizchi and M. K. Rafsanjani, "A survey on security challenges in cloud computing: issues, threats, and solutions," The journal of supercomputing, vol. 76, no. 12, pp. 9493-9532, 2020.

[107]J. Ni, Y. Yu, Y. Mu, and Q. Xia, "On the security of an efficient dynamic auditing protocol in cloud storage," IEEE Transactions on Parallel Distributed Systems, vol. 25, no. 10, pp. 2760-2761, 2014.

[108]N. Dhakad and J. Kar, "EPPDP: An Efficient Privacy-Preserving Data Possession With Provable Security in Cloud Storage," IEEE Systems Journal, 2022.

[109]C. Li and Z. Liu, "A Secure Privacy-Preserving Cloud Auditing Scheme with Data Deduplication," Int. J. Netw. Secur., vol. 21, no. 2, pp. 199-210, 2019.

[110]E. Daniel and N. Vasanthi, "LDAP: a lightweight deduplication and auditing protocol for secure data storage in cloud environment," Cluster Computing, vol. 22, no. 1, pp. 1247-1258, 2019.

[111]J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," presented at the 2013 IEEE Conference on Communications and Network Security (CNS), 2013.

[112]C. Li, J. Hu, K. Zhou, Y. Wang, and H. Deng, "Using blockchain for data auditing in cloud storage," in International Conference on Cloud Computing and Security, 2018: Springer, pp. 335-345.

[113]J. Xue, C. Xu, J. Zhao, and J. Ma, "Identity-based public auditing for cloud storage systems against malicious auditors via blockchain," Science China Information Sciences, vol. 62, no. 3, pp. 1-16, 2019.

[114] Y. Qi and Y. Huang, "DIRA: Enabling decentralized data integrity and reputation audit via blockchain," Sci. China Technological Sci, vol. 62, pp. 698-701, 2019.

[115] H. Wang, Z. Wang, and J. Domingo-Ferrer, "Anonymous and secure aggregation scheme in fog-based public cloud computing," Future Generation Computer Systems, vol. 78, pp. 712-719, 2018.

[116] D. Liu, Z. Yan, W. Ding, and M. Atiquzzaman, "A survey on secure data analytics in edge computing," IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4946-4967, 2019.

[117] B. Cao, L. Zhang, Y. Li, D. Feng, and W. Cao, "Intelligent offloading in multi-access edge computing: A state-of-the-art review and framework," IEEE Communications Magazine, vol. 57, no. 3, pp. 56-62, 2019.

[118] M. Dawood, S. Tu, C. Xiao, H. Alasmary, M. Waqas, and S. U. Rehman, “Cyberattacks and Security of Cloud Computing: A Complete Guideline,” Symmetry, vol. 15, no. 11, p. 1981, Oct. 2023, doi: 10.3390/sym15111981.

Promises, Challenges and Opportunities of Integrating SDN and Blockchain with IoT Applications: A Survey

Loubna Elhaloui[1], Mohamed Tabaa[2], Sanaa Elfilali[3], El habib Benlahmar[4]
Pluridisciplinary Laboratory of Research and Innovation (LPRI), EMSI Casablanca, Casablanca, Morocco[1, 2]
Laboratory of Information Technologies and Modelling-Faculty of Sciences Ben M’sik,
Hassan II University, Casablanca, Morocco[1, 3, 4]

***Abstract*—Security is a major issue in the IT world, and its aim is to maintain user confidence and the coherence of the entire information system. Various international and European research projects, as well as IT manufacturers, have proposed new solutions and mechanisms to solve the problem of security in the IoT environment. Software-Defined Networking (SDN) and Blockchain are advanced technologies utilized globally for establishing secure network communication and constructing resilient network infrastructures. They serve as a robust and dependable foundation for addressing various challenges, including security, privacy, scalability, and access control. Indeed, SDN and Blockchain technologies have demonstrated their ability to efficiently manage resource utilization and facilitate secure network communication within the Internet of Things (IoT) ecosystem. Nonetheless, there exists a research gap concerning the creation of a comprehensive framework that can fulfill the unique requirements of the IoT environment. Consequently, this paper presents a recent investigation into the integration of SDN and Blockchain with IoT. The objective is to analyze their primary contributions and identify the challenges involved. Subsequently, we offer relevant recommendations to address these challenges and enhance the security and privacy of the IoT landscape.**

***Keywords*—Internet of things; SDN; blockchain**

I. Introduction

The Internet of Things (IoT) paradigm is shaping the future of computing, rapidly integrating into our daily lives to enhance our quality of life by connecting various smart devices, technologies, services, and applications [1]. Nonetheless, managing IoT networks poses several challenges that demand innovative solutions. These challenges primarily revolve around the inherent vulnerabilities of IoT devices, including susceptibility to outages, failures under heavy traffic loads, security vulnerabilities, limited energy efficiency, and scalability issues. Given the heterogeneity and resource constraints of IoT devices, they require specialized network behaviors and services, such as security protocols, efficient power management, and load balancing modules.

Software-Defined Networking (SDN), with its novel network management approaches and recent advancements in the IoT domain, offers promising solutions. It grants global visibility into network status and enables logically centralized resource control, which can be physically distributed as needed via programmable APIs from a central point [2]. Consequently, SDN facilitates the implementation of innovative network management techniques. Consequently, considerable research efforts are dedicated to developing SDN-based IoT management frameworks [3]. While SDN lays the groundwork for robust management solutions, the integration of intelligent decision-making in uncertain scenarios is still lacking, necessitating the incorporation of AI-based approaches alongside SDN.

Blockchain and IoT are both innovations that can bring significant advantages to IoT networks, such as improved security, transparency, immutability, privacy, and automated business processes. However, when combined within an SDN framework for IoT network management, the potential benefits of these technologies are further amplified. Looking ahead, we envision the introduction of adaptive resource management frameworks for IoT networks with the assistance of AI, which will also incorporate blockchain-based SDN frameworks. Furthermore, as IoT is expected to undergo large-scale deployment in the future, practical challenges that go beyond controlled laboratory settings or theoretical assessments will emerge. The current state of research in this field suggests that the dynamic capabilities offered by SDN can be leveraged to reconfigure, update, and enhance IoT networks in real-time to address emerging challenges.

Despite the many advantages mentioned above, the integration of new emerging technologies into the security of information and communication systems can give rise to a number of problems. Critical situations, in particular, raise further questions. Since the integration of Blockchain and SDN into the Internet of Things (IoT) is a dynamic process influenced by several interdependent factors, the addition of Blockchain to the IoT ecosystem intensifies technological and organizational requirements. As such, this paper aims to present an in-depth study of the challenges associated with IoT integration. Consequently, attention is specifically focused on the following research questions:

RQ1: What current challenges hinder the integration of Blockchain and SDN technologies into the IoT?

RQ2: What guidance does the literature provide to surmount these challenges?

This paper offers several significant contributions. Firstly, despite the use of Blockchain in recent years, there is a lack of in-depth research into the challenges of the Internet of Things (IoT). This study proposes a comparative study of IoT security solutions, based on existing literature relating to its integration. The paper explores the current problems of decentralizing the IoT with Blockchain, as well as future challenges in this field.

The paper's structure is outlined as follows: In Section II, we delve into related work that is pertinent to our paper. Following that, in Section III, we present a comprehensive comparative analysis of technological solutions aimed at enhancing IoT security. Section IV is dedicated to the discussion of our findings and the presentation of results. Ultimately, the paper wraps up with a conclusion in Section V.

II. Related Work

A. Internet of Things with SDN

The authors in [4] examined the Manufacturer's Use Description (MUD) model, which encompasses network access control, data privacy, as well as policies for channel and authorization protection. They employed an SDN platform to efficiently access device data and resources and utilized Blockchain technology, specifically Hyperledger, for sharing data among IoT devices. Additionally, Molina and colleagues in [5] introduced a security framework for the continuous, on-demand management of virtual authentication, authorization, and accounting (AAA) in SDN-enabled IoT networks. Their work achieved scalable bootstrapping of IoT devices and fine-grained management of network access control.

Moreover, the authors in [6] introduced a novel combination of cloud computing, IoT, and SDN, resulting in the CENSOR framework. This framework was designed to establish a secure gateway within the IoT environment, featuring a reliable and secure IoT network architecture powered by cloud computing and based on SDN technology. The authors also highlighted several challenges and potential threats that need to be addressed, including advanced security measures to counter Distributed Denial of Ser-vice (DDoS) attacks, appropriate routing algorithms, and network scalability.

The authors in [7] proposed the use of a distributed controller cluster to address is-sues related to reliability, scalability, fault tolerance, and interoperability in an SDN network. Their method was found to maintain reasonable CPU utilization, thus optimizing controller performance, with a particular focus on enhancing the security of IoT applications.

Lastly, the authors in [8] introduced Middlebox-Guard (M-G), an SDN-based model for enhancing data transfer security in response to various attacks and to improve network stability. They first addressed the placement of middleboxes, which are associated with predefined security policies, using a placement selection algorithm. Subsequently, two SDN resource control algorithms were employed to fulfill coverage requirements within switching volume constraints. Simulation results demonstrated that the M-G model they designed could effectively enhance the security and stability of IoT networks.

B. Internet of Things with Blockchain

Cryptocurrency and financial transactions initially introduced blockchain technology, wherein all nodes within the blockchain execute and store all transactions. Blockchain, with its versatility, finds application in various domains, one of the most prominent being the Internet of Things (IoT) [9]. The IoT consists of networks of intelligent devices like Raspberry Pi, ESP, etc., which interconnect seamlessly to form a network used for sensing, processing, and communication. These smart IoT devices operate autonomously, consuming minimal energy and possessing lightweight characteristics. According to Statista Com [10], the estimated number of IoT devices in 2020 stood at 31 billion globally, and this number is projected to reach 75 billion devices by the end of 2025 [11].

In the IoT context, smart devices are primarily dedicated to energy-intensive tasks related to vital applications, making the implementation of privacy and security measures challenging. Malicious attacks can disrupt IoT services and pose threats to user privacy, data security, and overall network confidentiality [12]. These attacks in IoT-based systems fall into four main categories: physical attack, network attack, soft-ware attack, and data attack [13].

Physical Attack: In this category, attackers are physically close to the network and attempt to carry out malicious activities through various means, such as manipulating IoT devices, blocking RF signals, injecting malicious code, and conducting side-channel attacks. One countermeasure involves the use of physical non-clonable functions (PUF) to authenticate IoT devices, as it prevents physical attacks [14]. PUFs have a unique feature that makes it impossible to replicate the precise microstructure of an IoT device.

Network Attack: Attackers in this category seek to manipulate the IoT network through methods like RFID spoofing, man-in-the-middle attacks, traffic analysis, and Sybil attacks. Preventative measures include authentication techniques and secure hash functions [15].

Software Attack: Attackers exploit vulnerabilities in the current software of IoT systems.

Data Attack: This category involves unauthorized data access and data inconsistency. To thwart such attacks, blockchain technology can be used to provide effective privacy-preserving mechanisms [16].

Furthermore, several studies have proposed solutions to enhance IoT security:

In study [17], an algorithm secures the externalizations of bilinear pairings of IoT devices, significantly improving performance compared to standard bilinear matching.

The authors in [18] presents a framework based on a hybrid blockchain approach to secure multinational-level Industrial Internet of Things (IIoT) deployments across multiple countries.

In research [19], a patient-centric blockchain framework addresses data protection, authentication, and immutability concerns, built on the Hyperledger platform.

In study [20] proposes an Ethereum-based smart contract platform for a blockchain-driven healthcare infrastructure, offering potential efficiency improvements in hospital settings.

Lastly, in study [21] introduces an approach utilizing an open-source blockchain (Ethereum) to secure the outsourcing of bilinear pairs in IoT systems, addressing limitations of centralization and validating its effectiveness in securing IoT applications.

C. *Blockchain and SDN for the Internet of Things*

This section addresses the two research questions (RQ1: What current challenges hinder the integration of Blockchain and SDN technologies into the IoT? RQ2: What guidance does the literature provide to surmount these challenges?) We consulted the relevant literature for answers to the research questions. The search strategy encompassed the use of all well-known databases, including ACM digital library, Elsevier, IEEE and MDPI.

In their study [22], the authors introduced an innovative IoT architecture that effectively merges cutting-edge SDN and Blockchain technologies. The primary objective of this architecture is to integrate SDN controllers into IoT networks, utilizing a clustered structure along with a novel routing protocol to tackle various network challenges including security, privacy, access control, and availability. Their particular focus lies in devising energy-efficient mechanisms for data transfer among IoT devices within the SDN framework. This architecture harnesses both public and private blockchains to facilitate peer-to-peer communication between IoT devices and SDN controllers, incorporating a distributed trust authentication method.

Similarly, Chaudhary and colleagues [23] harnessed blockchain and SDN technologies to enhance the quality of service in an intelligent transportation system. They designed BEST, a blockchain-based secure energy exchange system for electric vehicles. BEST employs blockchain for decentralized validation of vehicle requests, thus eliminating single points of failure. Simulation results demonstrated the successful integration of blockchain into the SDN architecture, resulting in improved network QoS and more efficient energy usage, although it did not consider various energy sources.

The authors in [24] introduced Cochain-SC, a blockchain-based architecture that enables secure collaboration and decentralized attack information transfer among multiple SDN domains. This architecture combines intra-domain and inter-domain DDoS mitigation. The authors evaluated Cochain-SC's performance in terms of efficiency, security, cost-effectiveness, and the accuracy of detecting illegitimate flows.

Ferrag et al. [25] provided an extensive overview of Blockchain technology applications across various IoT domains, such as the Internet of Vehicles, the Internet of Energy, virtual web, cloud computing, and edge computing. Their study also addressed the five most common attacks in IoT networks, namely identity-based, cryptanalysis-based, reputation-based, manipulation-based, and service-based attacks. They established taxonomy of recent methods for achieving secure, privacy-preserving Blockchain technologies, comparing them based on specific models, security objectives, performance, computational complexity, limitations, and communication costs.

In study [26], the authors introduced the "DistBlockNet" framework, designed for a secure distributed SDN architecture for IoT through the use of Blockchain technology. They presented a scheme for updating and validating rule tables using Blockchain, with experimental evaluation demonstrating the effectiveness of DistBlockNet in terms of accuracy, scalability, defense capabilities, and performance overhead.

Blockchain technology, despite relying on a group of nodes that may not all be fully trustworthy, offers a dependable data structure thanks to its appropriate consensus algorithm. This makes it a valuable solution to address security challenges in IoT and SDN. In [27], decentralized security architecture based on SDN and block-chain for the IoT ecosystem was proposed, aiming to enhance attack detection and mitigation. Blockchain was employed for dynamic attack detection model updates and Fog node rewards based on proof-of-work.

Additionally, the authors in [28] introduced a blockchain-based controller designed to combat the injection of fake flow rules, primarily focusing on SDN controller authentication. The authors in [29] presented a novel blockchain-based authentication handover for an SDN-based 5G network, aiming to eliminate unnecessary reauthentication during repeated handovers between heterogeneous cells in 5G net-works.

Lastly, Qiu et al. [30] explored the Industrial Internet of Things scenario involving multiple SDN controllers. They proposed a blockchain-based consensus protocol for collecting and synchronizing network-wide views between different SDN controllers, employing the Q-learning method to optimize view switching, access selection, and computational resources.

In summary, various studies have proposed diverse solutions to integrate block-chain within an SDN-based IoT ecosystem. However, comprehensive scenarios that encompass all aspects of these works are yet to be fully developed.

III. Comparative Study of IoT Security Technology Solutions

The significance of combining blockchain and SDN with IoT has found application across a diverse range of domains. Table I highlights prevalent areas and recent research endeavors where the integration of blockchain and SDN with IoT applications plays a pivotal role. The majority of these studies are dedicated to enhancing security and privacy within the IoT landscape through the utilization of these two technologies. Notably, Table I reveals that these articles contend with specific challenges, notably in the realms of privacy and scalability.

TABLE I. RECENT STUDIES ON INTEGRATING BLOCKCHAIN AND SDN INTO THE IOT APPLICATIONS

Recent Survey Article	Years	Domain	IoT security	IoT privacy	Scalability	SDN-IoT	Blockchain
Oualha et al. [31]	2016	IoT device	√		√		
Mao et al. [32]	2016	IoT device	√				
Tonyali et al. [33]	2016	Smart Grid		√			
Hardjono et al. [34]	2016	IoT device	√		√		√
Hashemi et al. [35]	2016	IoT environment	√		√		√
Kokoris-K et al. [36]	2016	IoT device			√		√
Kamanashis et al. [37]	2016	Smart Cities	√		√		√
Bull et al. [38]	2016	IoT device	√			√	
Vandana et al. [39]	2016	IoT environment	√		√	√	
Gonzalez et al. [40]	2016	IoT environment	√		√	√	
Huh et al. [41]	2017	IoT device	√				√
Zhang and Wen [42]	2017	IoT E-business					√
Atlam et al. [43]	2020	IoT device					√
Khan and Salah [44]	2018	IoT System	√		√		√
Badr et al. [45]	2018	E-HEALTH	√	√			√
Uddin et al. [46]	2021	Cloud & Fog IoT					√
Dorri et al. [47]	2016	IoT System	√	√			√
Salman et al. [48]	2019	Cloud Computing	√	√			√
Dai et al. [49]	2019	Smart Industry	√				√
Zhang et al. [50]	2019	IoT System	√				
Lao et al. [51]	2021	IoT application					√
Patil et al. [52]	2018	Smart green house	√	√			√
Polyzos and Fotiou [53]	2017	IoT device	√				√
Zhu and Badr [54]	2018	IoT environment	√	√			
Mishra and Tyagi [55]	2019	E-HEALTH	√				√
Banerjee et al. [56]	2018	IoT SYSTEM	√				√
Wang et al. [57]	2019	IoT application	√				√
Sengupta et al. [58]	2020	Industrial IoT	√				√
Jesus et al. [59]	2018	IoT device	√	√			√
Dwivedi et al. [60]	2021	Industrial IoT					√
Kamilaris et al. [61]	2019	Smart Agriculture					√
Ferrag et al. [62]	2019	IoT Environment	√				
Zheng et al. [63]	2017	APPLICATION					√
Thakore et al. [64]	2019	IoT SYSTEM					√
Hassan et al. [65]	2019	IoT Application					√
Lin et al. [66]	2018	Smart Agriculture					√
Dogo et al. [67]	2019	Smart Agriculture	√				√
Kadam and John [68]	2020	IoT device	√	√			√
Maroufi et al. [69]	2019	IoT device					√
Alamri et al. [70]	2019	APPLICATION	√	√			√
Atlam and Wills [71]	2019	IoT device	√	√			
Saad et al. [72]	2019	IoT SYSTEM					√
Atlam et al. [73]	2019	IoT Environment	√				√
Tandon [74]	2019	APPLICATION	√	√			√
Karthikeyan et al. [75]	2019	APPLICATION		√			√
Fotiou et al. [76]	2018	Smart device	√	√			
Hang and Kim [77]	2019	IoT System	√	√			√
Mahmood et al. [78]	2021	IoT device	√				
R. Alcarria et al. [79]	2018	Smart Communities	√	√			√

Recent Survey Article	Years	Domain	IoT security	IoT privacy	Scalability	SDN-IoT	Blockchain
Oualha et al. [31]	2016	IoT device	√		√		
A. G. Ghandour [80]	2019	Smart City	√	√			√
A. Rahman [81]	2020	Smart Building	√	√	√	√	√
P. K. Sharma and J. H. Park [82]	2018	Smart City	√	√	√	√	√
Y. Gu, D. Hou [83]	2018	Cloud Computing	√	√			√
A. Yazdinejad [84]	2020	IoT Network	√	√	√	√	√
P. Singh [85]	2020	Smart City	√				√
P. K. Sharma [86]	2019	Smart Industry & Smart City	√	√	√		√
D. Sinh [87]	2018	Smart Network	√			√	
M. J. Islam [88]	2019	Smart City	√			√	
I. Abdulqadder [89]	2018	Cloud Environment				√	
R. Chaudhary [90]	2019	Smart Grid	√			√	√
P. K. Sharma [91]	2017	IoT Network	√			√	√
B. K. Mukherjee [92]	2020	Smart City	√			√	
A. Rahman [93]	2020	Smart Industry	√		√	√	√
A. Rahman [94]	2019	Smart City	√		√	√	√
Ali et al. [95]	2018	IoT Network	√	√	√	√	√
Alladi et al. [96]	2019	Smart Industry	√	√	√		√
Xie et al. [97]	2019	Smart City	√	√			√
Yang et al. [98]	2019	Edge Computing	√	√	√		√
Ahmed et al. [99]	2020	Smart City	√	√			√
Ferrag et al. [100]	2020	Smart Green agriculture	√		√		√
Wang et al. [101]	2020	Industrial IoT	√	√			√
Bhushan et al. [102]	2021	IoT Network	√		√		√
Majeed et al. [103]	2021	Smart City	√	√	√		√
Da Xu et al. [104]	2021	Edge Computing	√				√
Yaqoob et al. [105]	2021	E-health	√		√		√
Abdelmaboud et al. [106]	2022	IoT Application	√		√		√
Kumar et al. [107]	2022	Smart Industry	√	√			√
Yu et al. [108]	2022	Smart City	√				√
Pennino et al. [109]	2022	IoT Economy			√		√
Alkhateeb et al. [110]	2022	Hypride BC for IoT			√		

IV. Discussion

After examining 80 articles, it became evident that 80% of them were centered on enhancing the security of the Internet of Things (IoT). Additionally, 37% of these articles concentrated on matters pertaining to IoT privacy, while 30% delved into the evolution of IoT security. Consequently, blockchain has emerged as a potent and actively utilized tool for delivering the proposed security services for IoT applications.

Our initial approach involved categorizing diverse IoT applications by identifying their specific security requirements and the challenges they inherently face. Subsequently, we explored IoT solutions that addressed aspects of confidentiality, privacy, and availability, drawing upon conventional methods. Furthermore, we introduced emerging technologies such as Software Defined Networking (SDN) and Blockchain, both recognized for their effectiveness in mitigating scalability issues within the IoT ecosystem.

Moreover, we dedicated attention to security solutions that take into account the contextual aspects inherent to IoT applications, as delineated in Table II. We also considered the varying impacts of security concerns on system safety, alongside the corresponding countermeasures. Throughout our exploration, we provided an extensive comparative analysis of the different approaches, grounded in specific criteria. We also conducted an examination of techniques suitable for different types of IoT applications.

Despite ongoing efforts to confront the myriad challenges confronting the Internet of Things, numerous issues remain unresolved, notably those related to scalability and dynamism. This is particularly pertinent as the IoT continues its evolution into an "Internet of Everything," where humans, data, processes, and objects coalesce within a highly dynamic and intricately interconnected system.

TABLE II. NUMBER OF RELATED ARTICLES RANKED BY SCOPE OF CONTRIBUTION

Scope of the literature	Number of articles
IoT security	62
IoT privacy	29
Scalability	24
SDN-IoT	16
Blockchain	63
IoT security with Blockchain	48
IoT privacy with Blockchain	25
IoT security with SDN and Blockchain	9

V. CONCLUSION

The Internet of Things (IoT) confronts an array of security issues that surpass the complexities encountered in other domains, primarily owing to its intricate ecosystem and the inherent limitations of resource-constrained IoT devices. In recent years, an extensive body of research has been dedicated to addressing the diverse security challenges intimately linked with the IoT landscape. These challenges encompass intricate aspects such as authentication, confidentiality, integrity, access control, and policy enforcement, among a multitude of others.

Predominantly, prior works in the literature have strived to adapt security solutions initially devised for wireless and Internet sensor networks to suit the specific demands of the IoT framework. Nonetheless, it is imperative to underscore that IoT challenges present a novel dimension that proves considerably arduous to surmount using conventional remedies. Furthermore, it is crucial to emphasize that a significant proportion of prevailing security methodologies are rooted in centralized architectures, rendering their application within the IoT context notably more intricate due to the sheer abundance of interconnected entities. Consequently, a shift towards distributed approaches is imperative to effectively address the multifaceted security challenges that the IoT inherently presents.

REFERENCES

[1] Burhan, Muhammad, Rehman, Rana Asif, Khan, Bilal, et al. IoT elements, layered architectures and security issues: A comprehensive survey. sensors, 2018, vol. 18, no 9, p. 2796.

[2] Elhaloui, Loubna, Tabaa, Mohammed, Elfilali, Sanaa, et al. Dynamic security of IoT network traffic using SDN. Procedia Computer Science, 2023, vol. 220, p. 356-363.

[3] Siddiqui, Shahbaz, Hameed, Sufian, Shah, Syed Attique, et al. Towards Software-Defined-Networking-based IoT: A Systematic Literature Review on Management Frameworks and Open Challenges. 2021.

[4] S. N. Matheu, A. R. Enciso, A. M. Zarca, D. Garcia-Carrillo, J. L. Hernández-Ramos, J. B. Bernabe, and A. F. Skarmeta, ``Security architecture for de_ning and enforcing security pro_les in DLT/SDN-based IoT systems,'' Sensors, vol. 20, no. 7, p. 1882, Mar. 2020.

[5] A. M. Zarca, D. Garcia-Carrillo, J. B. Bernabe, J. Ortiz, R. Marin-Perez, and A. Skarmeta, ``Enabling virtual AAA management in SDN-based IoT networks,'' Sensors, vol. 19, no. 2, p. 295, Jan. 2019.

[6] M. Conti, P. Kaliyar, and C. Lal, ``CENSOR: Cloud-enabled secure IoT architecture over SDN paradigm,'' Concurrency Comput., Pract. Exper., vol. 31, no. 8, p. e4978, Apr. 2019.

[7] A. Abdelaziz, A. T. Fong, A. Gani, U. Garba, S. Khan, A. Akhunzada, H. Talebian, and K.-K.-R. Choo, ``Distributed controller clustering in software de_ned networks,'' PLoS ONE, vol. 12, no. 4, Apr. 2017, Art. no. e0174715.

[8] Y. Liu, Y. Kuang, Y. Xiao, and G. Xu, ``SDN-based data transfer security for Internet of Things,'' IEEE Internet Things J., vol. 5, no. 1, pp. 257_268, Feb. 2018.

[9] H. F. Atlam and G. B. Wills, "IoT security, privacy, safety and ethics," in Internet of Things, Springer International Publishing, 2020, pp. 123–149.

[10] A. Panarello, N. Tapas, G. Merlino, F. Longo, and A. Puliafito, "Blockchain and IoT integration: a systematic survey," Sensors, vol. 18, no. 8, Aug. 2018, doi: 10.3390/s18082575.

[11] M. H. Rehman, I. Yaqoob, K. Salah, M. Imran, P. P. Jayaraman, and C. Perera, "The role of big data analytics in industrial internet of things," Future Generation Computer Systems, vol. 99, pp. 247–259, Oct. 2019, doi: 10.1016/j.future.2019.04.020.

[12] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "A survey on the adoption of blockchain in IoT: challenges and solutionss," Blockchain: Research and Applications, vol. 2, no. 2, 2021, doi: 10.1016/j.bcra.2021.100006.

[13] J. Sengupta, S. Ruj, and S. Das Bit, "A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoTt," Journal of Network and Computer Applications, vol. 149, 2020, doi: 10.1016/j.jnca.2019.102481.

[14] K. Mahmood et al., "PUF enable lightweight key-exchange and mutual authentication protocol for multi-server based D2D communication," Journal of Information Security and Applications, vol. 61, Sep. 2021, doi: 10.1016/j.jisa.2021.102900.

[15] D. Mishra, P. Vijayakumar, V. Sureshkumar, R. Amin, S. H. Islam, and P. Gope, "Efficient authentication protocol for secure multimedia communications in IoT-enabled wireless sensor networks," Multimedia Tools and Applications, vol. 77, no. 14, pp. 18295–18325, Nov. 2018, doi: 10.1007/s11042-017-5376-4.

[16] S. K. Dwivedi, R. Amin, and S. Vollala, "Blockchain-based secured IPFS-enable event storage technique with authentication protocol in VANET," IEEE/CAA Journal of Automatica Sinica, vol. 8, no. 12, pp. 1913–1922, Dec. 2021, doi: 10.1109/JAS.2021.1004225.

[17] Zhang, H.; Tong, L.; Yu, J.; Lin, J. Blockchain Aided Privacy-Preserving Outsourcing Algorithms of Bilinear Pairings for Internet of Things Devices. arXiv 2021, arXiv:2101.02341.

[18] Rathee, G.; Ahmad, F.; Sandhu, R.; Kerrache, C.A.; Azad, M.A. On the design and implementation of a secure blockchain-based hybrid framework for Industrial Internet-of-Things. Inf. Process. Manag. 2021, 58, 102526.

[19] Singh, A.P.; Pradhan, N.R.; Agnihotri, S.; Jhanjhi, N.; Verma, S.; Ghosh, U.; Roy, D. A Novel Patient-Centric Architectural Framework for Blockchain-Enabled Healthcare Applications. IEEE Trans. Ind. Inform. 2020, 17, 5779–5789.

[20] Latif, R.M.A.; Hussain, K.; Jhanjhi, N.; Nayyar, A.; Rizwan, O. A remix IDE: Smart contract-based framework for the healthcare sector by using Blockchain technology. Multimed. Tools Appl. 2020, 1–24.

[21] Lin, C.; He, D.; Huang, X.; Xie, X.; Choo, K.-K.R. Blockchain-based system for secure outsourcing of bilinear pairings. Inf. Sci. 2020, 527, 590–601.

[22] A. Yazdinejad, R. M. Parizi, A. Dehghantanha, Q. Zhang, and K.-K.-R. Choo, ``An energy-ef_cient SDN controller architecture for IoT networks with blockchain-based security,'' IEEE Trans. Services Comput., vol. 13, no. 4, pp. 625_638, Jul. 2020.

[23] R. Chaudhary, A. Jindal, G. S. Aujla, S. Aggarwal, N. Kumar, and K.-K.-R. Choo, ``BEST: Blockchain-based secure energy trading in SDNenabled intelligent transportation system,'' Comput. Secur., vol. 85, pp. 288_299, Aug. 2019.

[24] Z. A. El Houda, A. S. Ha_d, and L. Khoukhi, ``Cochain-SC: An intra- and inter-domain ddos mitigation scheme based on blockchain using SDN and smart contract,'' IEEE Access, vol. 7, pp. 98893_98907, 2019.

[25] M. A. Ferrag, M. Derdour, M. Mukherjee, A. Derhab, L. Maglaras, and H. Janicke, ``Blockchain technologies for the Internet of Things:

Research issues and challenges," IEEE Internet Things J., vol. 6, no. 2, pp. 2188_2204, Apr. 2019.

[26] P. K. Sharma, S. Singh, Y.-S. Jeong, and J. H. Park, ``DistBlockNet: A distributed blockchains-based secure SDN architecture for IoT networks," IEEE Commun. Mag., vol. 55, no. 9, pp. 78_85, Sep. 2017.

[27] S. Rathore, B. W. Kwon, and J. H. Park, "Blockseciotnet: Blockchainbased decentralized security architecture for iot network," Journal of Network and Computer Applications, vol. 143, pp. 167–177, 2019.

[28] S. Boukria, M. Guerroumi, and I. Romdhani, "BCFR: Blockchain-based controller against false flow rule injection in SDN," in 2019 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2019, pp. 1034–1039.

[29] A. Yazdinejad, R. M. Parizi, A. Dehghantanha, and K.-K. R. Choo, "Blockchain-enabled authentication handover with efficient privacy protection in sdn-based 5g networks," IEEE Transactions on Network Science and Engineering, 2019.

[30] C. Qiu, F. R. Yu, H. Yao, C. Jiang, F. Xu, and C. Zhao, "Blockchainbased software-defined industrial internet of things: A dueling deep qlearning approach," IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4627–4639, 2018.

[31] N. Oualha and K. T. Nguyen. Lightweight attribute-based encryption for the internet of things. In 2016 25th International Conference on Computer Communication and Networks (ICCCN), pages 1–6. IEEE, 2016.

[32] Y. Mao, J. Li, M.-R. Chen, J. Liu, C. Xie, and Y. Zhan. Fully secure fuzzy identity-based encryption for secure iot communications. Computer Standards & Interfaces, 44 :117–121, 2016.

[33] S. Tonyali, O. Cakmak, K. Akkaya, M. M. Mahmoud, and I. Guvenc. Secure data obfuscation scheme to enable privacy-preserving state estimation in smart grid ami networks. IEEE Internet of Things Journal, 3(5) :709–719, 2016.

[34] T. Hardjono and N. Smith. Cloud-based commissioning of constrained devices using permissioned blockchains. In Proceedings of the 2nd ACM International Workshop on IoT Privacy, Trust, and Security, pages 29– 36. ACM, 2016.

[35] S. H. Hashemi, F. Faghri, P. Rausch, and R. H. Campbell. World of empowered iot users. In 2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI), pages 13–24. IEEE, April 2016.

[36] L. Kokoris-Kogias, L. Gasser, I. Khoffi, P. Jovanovic, N. Gailly, and B. Ford. Managing identities using blockchains and cosi. In 9th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2016), number EPFL-TALK-220210, 2016.

[37] K. Biswas and V. Muthukkumarasamy. Securing smart cities using blockchain technology. In 2016 IEEE 18th International Conference on High Performance Computing and communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pages 1392–1393. IEEE, Dec 2016.

[38] P. Bull, R. Austin, E. Popov, M. Sharma, and R. Watson. Flow based security for iot devices using an sdn gateway. In 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Future Internet of Things and Cloud (FiCloud),, pages 157–163. IEEE, July 2016.

[39] C. Vandana. Security improvement in iot based on software defined networking (sdn). International Journal of Engineering and Technology Research (IJSETR), 5(1) :291–295, january 2016.

[40] C. Gonzalez, S. M. Charfadine, O. Flauzac, and F. Nolot. Sdnbased security framework for the iot in distributed grid. In 2016 International Multidisciplinary Conference on Computer and Energy Science (SpliTech), pages 1–5. IEEE, July 2016.

[41] S. Huh, S. Cho, and S. Kim, "Managing IoT devices using blockchain platform," in International Conference on Advanced Communication Technology, 2017, pp. 464–467, doi: 10.23919/ICACT.2017.7890132.

[42] Y. Zhang and J. Wen, "The IoT electric business model: Using blockchain technology for the internet of things," Peer-to-Peer Networking and Applications, vol. 10, no. 4, pp. 983–994, Jul. 2017, doi: 10.1007/s12083-016-0456-1.

[43] H. F. Atlam, M. A. Azad, A. G. Alzahrani, and G. Wills, "A review of blockchain in internet of things and AI," Big Data and Cognitive Computing, vol. 4, no. 4, pp. 1–27, Oct. 2020, doi: 10.3390/bdcc4040028.

[44] M. A. Khan and K. Salah, "IoT security: review, blockchain solutions, and open challenges," Future Generation Computer Systems, vol. 82, pp. 395–411, May 2018, doi: 10.1016/j.future.2017.11.022.

[45] S. Badr, I. Gomaa, and E. Abd-Elrahman, "Multi-tier blockchain framework for IoT-EHRs systems," Procedia Computer Science, vol. 141, pp. 159–166, 2018, doi: 10.1016/j.procs.2018.10.162.

[46] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "A survey on the adoption of blockchain in IoT: challenges and solutionss," Blockchain: Research and Applications, vol. 2, no. 2, 2021, doi: 10.1016/j.bcra.2021.100006.

[47] A. Dorri, S. S. Kanhere, and R. Jurdak, "Blockchain in internet of things: challenges and solutions," arxiv.org/abs/1608.05187, 2016, [Online]. Available: http://arxiv.org/abs/1608.05187.

[48] T. Salman, M. Zolanvari, A. Erbad, R. Jain, and M. Samaka, "Security services using blockchains: a state-of-the-art survey," IEEE Communications Surveys and Tutorials, vol. 21, no. 1, pp. 858–880, 2019, doi: 10.1109/COMST.2018.2863956.

[49] H. N. Dai, Z. Zheng, and Y. Zhang, "Blockchain for internet of things: a survey," IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8076–8094, Oct. 2019, doi: 10.1109/JIOT.2019.2920987.

[50] Y. Zhang, S. Kasahara, Y. Shen, X. Jiang, and J. Wan, "Smart contract-based access control for the internet of things," IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1594–1605, Apr. 2019, doi: 10.1109/JIOT.2018.2847705.

[51] L. Lao, Z. Li, S. Hou, B. Xiao, S. Guo, and Y. Yang, "A survey of IoT applications in blockchain systems," ACM Computing Surveys, vol. 53, no. 1, pp. 1–32, Jan. 2021, doi: 10.1145/3372136.

[52] A. S. Patil, B. A. Tama, Y. Park, and K. H. Rhee, "A framework for blockchain based secure smart green house farming," in Lecture Notes in Electrical Engineering, vol. 474, Springer Singapore, 2018, pp. 1162–1167.

[53] G. C. Polyzos and N. Fotiou, "Blockchain-assisted information distribution for the internet of things," in IEEE International Conference on Information Reuse and Integration (IRI), Aug. 2017, pp. 75–78, doi: 10.1109/IRI.2017.83.

[54] X. Zhu and Y. Badr, "Identity management systems for the internet of things: a survey towards blockchain solutions," Sensors, vol. 18, no. 12, Dec. 2018, doi: 10.3390/s18124215.

[55] S. Mishra and A. K. Tyagi, "Intrusion detection in internet of things (IoTs) based applications using blockchain technology," in Proceedings of the 3rd International Conference on I-SMAC IoT in Social, Mobile, Analytics and Cloud, Dec. 2019, pp. 123–128, doi: 10.1109/I-SMAC47947.2019.9032557.

[56] M. Banerjee, J. Lee, and K.-K. R. Choo, "A blockchain future for internet of things security: a position paper," Digital Communications and Networks, vol. 4, no. 3, pp. 149–160, Aug. 2018, doi: 10.1016/j.dcan.2017.10.006.

[57] X. Wang et al., "Survey on blockchain for internet of things," Computer Communications, vol. 136, pp. 10–29, Feb. 2019, doi: 10.1016/j.comcom.2019.01.006.

[58] J. Sengupta, S. Ruj, and S. Das Bit, "A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoTt," Journal of Network and Computer Applications, vol. 149, 2020, doi: 10.1016/j.jnca.2019.102481.

[59] E. F. Jesus, V. R. L. Chicarino, C. V. N. De Albuquerque, and A. A. D. A. Rocha, "A survey of how to use blockchain to secure internet of things and the stalker attack," Security and Communication Networks, vol. 2018, pp. 1–27, Apr. 2018, doi: 10.1155/2018/9675050.

[60] S. K. Dwivedi, P. Roy, C. Karda, S. Agrawal, and R. Amin, "Blockchain-based internet of things and industrial IoT: a comprehensive survey," Security and Communication Networks, vol. 2021, pp. 1–21, Aug. 2021, doi: 10.1155/2021/7142048.

[61] A. Kamilaris, A. Fonts, and F. X. Prenafeta-Boldύ, "The rise of blockchain technology in agriculture and food supply chains," Trends in

Food Science and Technology, vol. 91, pp. 640–652, Sep. 2019, doi: 10.1016/j.tifs.2019.07.034.

[62] M. A. Ferrag, L. Maglaras, and H. Janicke, “Blockchain and its role in the internet of things,” in Springer Proceedings in Business and Economics, Springer International Publishing, 2019, pp. 1029–1038.

[63] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, “An overview of blockchain technology: architecture, consensus, and future trends,” in Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017, Jun. 2017, pp. 557–564, doi: 10.1109/BigDataCongress.2017.85.

[64] R. Thakore, R. Vaghashiya, C. Patel, and N. Doshi, “Blockchain - based IoT: a survey,” Procedia Computer Science, vol. 155, pp. 704–709, 2019, doi: 10.1016/j.procs.2019.08.101.

[65] F. Hassan et al., “Blockchain and the future of the internet: a comprehensive review,” arxiv.org/abs/1904.00733, Feb. 2019, [Online]. Available: http://arxiv.org/abs/1904.00733.

[66] J. Lin, Z. Shen, A. Zhang, and Y. Chai, “Blockchain and IoT based food traceability for smart agriculture,” in Proceedings of the 3rd International Conference on Crowd Science and Engineering, 2018, pp. 1–6, doi: 10.1145/3265689.3265692.

[67] E. M. Dogo, A. F. Salami, N. I. Nwulu, and C. O. Aigbavboa, “Blockchain and internet of things-based technologies for intelligent water management system,” in Artificial Intelligence in IoT, Springer International Publishing, 2019, pp. 129–150.

[68] S. B. Kadam and S. K. John, “Blockchain integration with low-power internet of things devices,” in Handbook of Research on Blockchain Technology, Elsevier, 2020, pp. 183–211.

[69] M. Maroufi, R. Abdolee, and B. M. Tazekand, “On the convergence of blockchain and internet of things (IoT) technologies,” Journal of Strategic Innovation and Sustainability, vol. 14, no. 1, Mar. 2019, doi: 10.33423/jsis.v14i1.990.

[70] M. Alamri, N. Z. Jhanjhi, and M. Humayun, “Blockchain for internet of things (IoT) research issues challenges & future directions: a review,” International Journalof Computer Science and Network Security, 2019.

[71] H. F. Atlam and G. B. Wills, “Intersections between IoT and distributed ledger,” in Advances in Computers, vol. 115, Elsevier, 2019, pp. 73–113.

[72] M. Saad et al., “Exploring the attack surface of blockchain: a systematic overview,” arxiv.org/abs/1904.03487, Apr. 2019, [Online]. Available: http://arxiv.org/abs/1904.03487.

[73] H. F. Atlam and G. B. Wills, “An efficient security risk estimation technique for risk-based access control model for IoT,” Internet of Things, vol. 6, Jun. 2019, doi: 10.1016/j.iot.2019.100052.

[74] A. Tandon, “An empirical analysis of using blockchain technology with internet of things and its application,” International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9S3, pp. 1469–1475, Aug. 2019, doi: 10.35940/ijitee.I3310.0789S319.

[75] P. Karthikeyyan, S. Velliangiri, and I. T. Joseph, “Review of blockchain based IoT application and its security issues,” in 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, Jul. 2019, pp. 6–11, doi: 10.1109/ICICICT46008.2019.8993124.

[76] N. Fotiou, V. A. Siris, and G. C. Polyzos, “Interacting with the internet of things using smart contracts and blockchain technologies,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11342, Springer International Publishing, 2018, pp. 443–452.

[77] L. Hang and D.-H. Kim, “Design and implementation of an integrated IoT blockchain platform for sensing data integrity,” Sensors, vol. 19, no. 10, May 2019, doi: 10.3390/s19102228.

[78] K. Mahmood et al., “PUF enable lightweight key-exchange and mutual authentication protocol for multi-server based D2D communication,” Journal of Information Security and Applications, vol. 61, Sep. 2021, doi: 10.1016/j.jisa.2021.102900.

[79] R. Alcarria, B. Bordel, T. Robles, D. Martín, and M.-Á. Manso-Callejo, ``A blockchain-based authorization system for trustworthy resource monitoring and trading in smart communities,'' Sensors, vol. 18, no. 10, p. 3561, Oct. 2018.

[80] A. G. Ghandour, M. Elhoseny, and A. E. Hassanien, ``Blockchains for smart cities: A survey,'' in Security in Smart Cities: Models, Applications, and Challenges. Springer, 2019, pp. 193_210.

[81] A. Rahman, M. K. Nasir, Z. Rahman, A. Mosavi, and B. Minaei-Bidgoli, ``DistBlockBuilding: A distributed blockchainbased SDN-IoT network for smart building management,'' IEEE Access, vol. 8, pp. 140008_140018, 2020.

[82] P. K. Sharma and J. H. Park, ``Blockchain based hybrid network architecture for the smart city,'' Future Gener. Comput. Syst., vol. 86, pp. 650_655, Sep. 2018.

[83] Y. Gu, D. Hou, X. Wu, J. Tao, and Y. Zhang, ``Decentralized transaction mechanism based on smart contract in distributed data storage,'' Information, vol. 9, no. 11, p. 286, Nov. 2018.

[84] A. Yazdinejad, R. M. Parizi, A. Dehghantanha, Q. Zhang, and K.-K.-R. Choo, ``An energy-ef_cient SDN controller architecture for IoT networks with blockchain-based security,'' IEEE Trans. Services Comput., vol. 13, no. 4, pp. 625_638, Jul. 2020.

[85] P. Singh, A. Nayyar, A. Kaur, and U. Ghosh, ``Blockchain and fog based architecture for Internet of everything in smart cities,'' Future Internet, vol. 12, no. 4, p. 61, Mar. 2020.

[86] P. K. Sharma, N. Kumar, and J. H. Park, ``Blockchain-based distributed framework for automotive industry in a smart city,'' IEEE Trans. Ind. Informat., vol. 15, no. 7, pp. 4197_4205, Jul. 2019.

[87] D. Sinh, L.-V. Le, B.-S. P. Lin, and L.-P. Tung, ``SDN/NFV_A new approach of deploying network infrastructure for IoT,'' in Proc. 27th Wireless Opt. Commun. Conf. (WOCC), Apr./May 2018, pp. 1_5.

[88] M. J. Islam, M. Mahin, S. Roy, B. C. Debnath, and A. Khatun, ``DistBlackNet: A distributed secure black SDN-IoT architecture with NFV implementation for smart cities,'' in Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE), Feb. 2019, pp. 1_6.

[89] I. Abdulqadder, D. Zou, I. Aziz, B. Yuan, and W. Dai, ``Deployment of robust security scheme in SDN based 5G network over NFV enabled cloud environment,'' IEEE Trans. Emerg. Topics Comput., early access, Nov. 5, 2018.

[90] R. Chaudhary, A. Jindal, G. S. Aujla, S. Aggarwal, N. Kumar, and K.-K.-R. Choo, ``BEST: Blockchain-based secure energy trading in SDN-enabled intelligent transportation system,'' Comput. Secur., vol. 85, pp. 288_299, Aug. 2019.

[91] P. K. Sharma, S. Singh, Y.-S. Jeong, and J. H. Park, ``DistBlockNet: A distributed blockchains-based secure SDN architecture for IoT networks,'' IEEE Commun. Mag., vol. 55, no. 9, pp. 78_85, Sep. 2017.

[92] B. K. Mukherjee, M. S. I. Pappu, M. J. Islam, and U. K. Acharjee, ``An SDN based distributed IoT network with NFV implementation for smart cities,'' in Proc. 2nd Int. Conf. Cyber Secur. Comput. Sci. (ICONCS). Springer, 2020, pp. 539_552.

[93] A. Rahman, U. Sara, D. Kundu, S. Islam, M. Jahidul, M. Hasan, Z. Rahman, and M. Kamal, ``DistB-SDoIndustry: Enhancing security in industry 4.0 services based on distributed blockchain through software de_ned networking-IoT enabled architecture,'' Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 9, 2020.

[94] A. Rahman, M. J. Islam, F. A. Sunny, and M. K. Nasir, ``DistblockSDN: A distributed secure blockchain based SDN-IoT architecture with NFV implementation for smart cities,'' in Proc. Int. Conf. Innov. Eng. Technol. (ICIET), 2019, pp. 23_24.

[95] Ali, M.S.; Vecchio, M.; Pincheira, M.; Dolui, K.; Antonelli, F.; Rehmani, M.H. Applications of blockchains in the internet of things: A comprehensive survey. IEEE Commun. Surv. Tutor. 2018, 21, 1676–1717.

[96] Alladi, T.; Chamola, V.; Parizi, R.M.; Choo, K.-K.R. Blockchain applications for industry 4.0 and industrial IoT: A review. IEEE Access 2019, 7, 176935–176951.

[97] Xie, J.; Tang, H.; Huang, T.; Yu, F.R.; Xie, R.; Liu, J.; Liu, Y. A survey of blockchain technology applied to smart cities: Research issues and challenges. IEEE Commun. Surv. Tutor. 2019, 21, 2794–2830.

[98] Yang, R.; Yu, F.R.; Si, P.; Yang, Z.; Zhang, Y. Integrated blockchain and edge computing systems: A survey, some research issues and challenges. IEEE Commun. Surv. Tutor. 2019, 21, 1508–1532.

[99] Ahmed, S.; Shah, M.A.; Wakil, K. Blockchain as a trust builder in the smart city domain: A systematic literature review. IEEE Access 2020, 8, 92977–92985.

[100] Ferrag, M.A.; Shu, L.; Yang, X.; Derhab, A.; Maglaras, L. Security and privacy for green IoT-based agriculture: Review, blockchain solutions, and challenges. IEEE Access 2020, 8, 32031–32053.

[101] Wang, Q.; Zhu, X.; Ni, Y.; Gu, L.; Zhu, H. Blockchain for the IoT and industrial IoT: A review. Internet Things 2020, 10, 100081.

[102] Bhushan, B.; Sahoo, C.; Sinha, P.; Khamparia, A. Unification of blockchain and internet of things (BIoT): Requirements, working model, challenges and future directions. Wirel. Netw. 2021, 27, 55–90.

[103] Majeed, U.; Khan, L.U.; Yaqoob, I.; Kazmi, S.A.; Salah, K.; Hong, C.S. Blockchain for IoT-based smart cities: Recent advances, requirements, and future challenges. J. Netw. Comput. Appl. 2021, 181, 103007.

[104] Da Xu, L.; Lu, Y.; Li, L. Embedding blockchain technology into IoT for security: A survey. IEEE Internet Things J. 2021, 8, 10452–10473.

[105] Yaqoob, I.; Salah, K.; Jayaraman, R.; Al-Hammadi, Y. Blockchain for healthcare data management: Opportunities, challenges, and future recommendations. Neural Comput. Appl. 2022, 34, 11475–11490.

[106] Abdelmaboud, A.; Ahmed, A.I.A.; Abaker, M.; Eisa, T.A.E.; Albasheer, H.; Ghorashi, S.A.; Karim, F.K. Blockchain for IoT Applications: Taxonomy, Platforms, Recent Advances, Challenges and Future Research Directions. Electronics 2022, 11, 630.

[107] Kumar, R.L.; Khan, F.; Kadry, S.; Rho, S. A Survey on blockchain for industrial Internet of Things. Alex. Eng. J. 2022, 61, 6001–6022.

[108] Yu, Z.; Song, L.; Jiang, L.; Sharafi, O.K. Systematic literature review on the security challenges of blockchain in IoT-based smart cities. Kybernetes 2021, 51.

[109] Pennino, D.; Pizzonia, M.; Vitaletti, A.; Zecchini, M. Blockchain as IoT Economy enabler: A review of architectural aspects. J. Sens. Actuator Netw. 2022, 11, 20.

[110] Alkhateeb, A.; Catal, C.; Kar, G.; Mishra, A. Hybrid blockchain platforms for the internet of things (IoT): A systematic literature review. Sensors 2022, 22, 1304.

Application of Machine Learning in Learning Problems and Disorders: A Systematic Review

Mario Aquino Cruz[1], Oscar Alcides Choquehuallpa Hurtado[2], Esther Calatayud Madariaga[3]
Departamento Académico de Informática y Sistemas, Universidad Nacional Micaela Bastidas de Apurímac, Abancay, Perú[1, 2]
Departamento Académico de Ciencias Básicas, Universidad Nacional Micaela Bastidas de Apurímac, Abancay, Perú[3]

Abstract—**Learning Disorders, which affect approximately 10% of the school population, represent a significant challenge in the educational field. The lack of proper diagnosis and treatment can have profound consequences, triggering psychological problems in those affected by disorders that impact reading, writing, numeracy and attention, among others. Notable among them are Attention Deficit Hyperactivity Disorder (ADHD) and dyslexia. In this context, a literature review focusing on Machine Learning applications to address these educational problems is addressed. The methodology proposed by Barbara Kitchenham guides this analysis, using the online tool Parsifal for the review, generation of search strings, formulation of research questions and management of information sources. The first findings of this research highlight a growing trend in the application of Machine Learning techniques in learning problems and disorders, especially in the last five years, as of 2019. Among the primary sources, the IEEE Digital Library emerges as a key source of information in this rapidly developing field. This innovative approach has the potential to significantly improve early detection, accurate diagnosis and implementation of personalized interventions, thus offering new perspectives in understanding and addressing the educational challenges associated with Learning Disorders.**

Keywords—Machine learning; learning disorder; deep learning; ADHD; dyslexia; learning impairment

I. Introduction

Learning Disorders (LD) are the most prevalent neurodevelopmental disorders in the population, affecting about 10% of the population at school age. AT cause that children with an adequate schooling and normal intelligence with difficulties of neurobiological origin, due to the lack of a proper diagnosis and treatment, suffer this problem of learning disorders causing psychological problems to those affected. These disorders affect reading, writing, calculation and/or attention, among others [1]. Some of the best known disorders are ADHD and dyslexia.

ADHD attention [2] deficit and hyperactivity disorder, considered as a specific learning disorder, as well as a behavioral disorder, has been one of the main reasons for consultation in pediatric neurology, although the clinical criteria are well established, probably the clinical history is not rigorously reviewed and the most determinant symptoms are not adequately noted.

Dyslexia is a brain condition that makes reading, spelling, writing and sometimes speaking difficult. The brains of people with dyslexia have difficulty recognizing or processing certain types of information. It occurs in people who do not have any physical, motor, visual or other disabilities. [3] Likewise, people with dyslexia have normal cognitive development.

Learning disorders are not considered problems, although they were for many years, but thanks to advances in neuroscience, it is now defined as a learning disorder, and the causes are neurobiological in origin. [3] The child is born and the cause is generally of genetic origin, although several institutions still maintain the name of learning disabilities, it must be clear that they are not problems but learning disorders.

It is therefore important to recognize the importance of approaching Learning Disorders from a holistic perspective. While science has advanced in understanding their neurobiological origin, it is essential to emphasize the need for early detection and a multidisciplinary approach to ensure accurate diagnosis and appropriate treatment. Lack of awareness and understanding of these disorders can lead to children facing unnecessary difficulties in their academic and emotional development. Therefore, advocating for greater awareness, education and support for those affected is key to building an inclusive society that recognizes and values diversity in learning abilities.

Machine Learning techniques can help in investigating and addressing learning disorders, such as Attention Deficit Hyperactivity Disorder (ADHD) and dyslexia, due to their unique ability to analyze vast amounts of heterogeneous data and extract complex patterns that may go unnoticed by conventional methods. These disorders, characterized by a diversity of clinical manifestations and associated factors, present significant challenges for accurate diagnosis and a thorough understanding of their underlying mechanisms. The application of these algorithms allows the integration of information from various sources, such as reading tests, magnetic resonance imaging and electroencephalography, thus facilitating a multidimensional approach. In addition, the classification and predictive capabilities of these techniques can be crucial to improve early detection, personalize intervention strategies and move towards the development of automated diagnostic systems.

The aim of this article is to make a systematic review on the contributions of Machine Learning and Deep learning in learning disabilities and learning disorders detection. For which a methodology inspired by Barbara Kitchenham's proposal was used, then in the results section the information extracted from the selected articles is analyzed and finally the conclusions obtained are specified, the information collected is intended to serve as a basis for other studies related to the application of Machine Learning in the problems of learning disorders. It is also intended to motivate readers to delve deeper into this area of research.

II. Methodology

For the development of this article, Barbara Kitchenham's methodology was applied as an inspiration, it is important to highlight that it helps significantly in works related to the systematic review of the literature, the phases that helped to develop this study are described below.

Plan the systematic literature review using the Parsifal tool.

- Determine research questions.
- Establish search process.
- Define inclusion and exclusion criteria for articles.
- Select query sources.
- Create search strings.

Develop the systematic literature review with the defined planning.

- Search for articles.
- Selection of definitive articles for information analysis.
- Analysis and classification of information.

Document and interpret the results of the review.

- Develop the report of the research questions of the review.

The tools used for the development of this research are as follows:

- Mendeley: It is a tool that allows the management and administration of bibliographic sources. It allowed the management of the references mentioned in this research work.
- Parsifal: It is a web tool that helped with the creation of search strings, keywords and research questions.

III. Development

A. Plan the Systematic Literature Review

Each of the tasks performed during the planning stage of the systematic literature review are detailed below:

1) Determine the research questions: Based on the purpose of this article, the following research questions are raised.

- In which repositories have most of the articles on the topic of study been published?
- In what year were most articles published, and how are they interpreted?
- How does Machine Learning help in the detection of learning problems and disorders?

2) Establish the search process: Base terms were identified by applying the PICOC method [4] to define the scope of the systematic review. Its components are population, intervention, comparison, results and contexts. This method made it possible to define the expressions that make up the search strings. They are detailed below:

- Population (P): “Learning disorder” OR “Learning Impairment” OR “ADHD” OR “Learning Deficit”.
- Intervention (I): “Machine learning” OR “Deep Learning”.
- Comparison (C): Not applicable.
- Outcomes (O): “Algorithms” OR “Classification” OR “Detection” OR “Methods” OR “Techniques”
- Context (C): “Artificial Intelligence”.

3) Define the inclusion and exclusion criteria for the articles: According to the objectives and scope of this article, it is necessary to establish four inclusion criteria (IC) and four exclusion criteria (EC), which are described as follows:

Inclusion criteria (IC):

- IC1: Articles containing information on Machine Learning or Deep Learning techniques.
- IC2: Articles written in English or Spanish.
- IC3: Articles published from 2019 onwards.
- IC4: Articles that have been published in scientific journals, scientific articles, and scientific articles.

Inclusion criteria (EC):

- EC1: Duplicate articles.
- EC2: Articles whose title is not related to the object of study.
- EC3: Articles that do not belong to the area of Science and Computing
- EC4: Book chapters, manuals, gray literature.

4) Select sources of consultation: Four search sources indicated in Table I were selected due to their accessibility and advanced query support.

TABLE I. Scientific Databases

Database	URL
IEEE	https://ieeexplore.ieee.org
ScienceDirect	https://www.sciencedirect.com
Scopus	https://www.scopus.com
Doaj	https://doaj.org/

5) Search Strings: The keywords were considered on the basis of what was applied in the PICOC method, and the logical operators "AND/OR" were used to generate the search strings.

The Parsifal tool generated the general search string, which was modified according to each database, and all the keywords were defined in English, as shown in Table II.

TABLE II. SEARCH STRING

Database	Search Strings
Parsifal	("Learning disorder" OR "ADHD" OR "Learning Deficit" OR "Learning Impairment" OR "TDAH") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques")
IEEE	(("Learning disorder" OR "ADHD" OR "Learning Deficit" OR "Learning Impairment" OR "TDAH") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques"))
ScienceDirect	(("Learning disorder" OR "ADHD") AND ("Machine Learning" AND "Deep Learning") AND ("Algorithms" AND "Classification" OR "Detection" AND ("Methods" OR "Techniques"))
Scopus	TITLE-ABS-KEY(("Learning disorder" OR "Learning Deficit" OR "Learning Impairment" OR "ADHD" OR "TDAH") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" AND ("Classification" OR "Detection") AND ("Methods" OR "Techniques"))) AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"cp")) AND (LIMIT-TO (SUBJAREA,"COMP")) AND (LIMIT-TO (PUBYEAR,2023) OR LIMIT-TO (PUBYEAR,2022) LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019))
Doaj	"Learning disorder" AND "machine learning"

B. Develop the Systematic Literature Review

1) Search for articles: The search string was implemented in each of the selected databases (IEEE, ScienceDirect, Scopus and Doaj), where 413 articles were obtained. Of these, articles that did not meet the inclusion criteria were ignored and duplicate articles or those that had no relevance to the field of study were discarded. Of the 413 articles, 22 were selected for further review as they met the prerequisites.

2) Selection of definitive items for data analysis: The norm in this section, according to the methodology proposed by Barbara Kitchenham, is that the articles go through a series of quality questions in order to select the articles that contribute the most to the research objective. However, since there are only 22 articles and all meet the requirements for inclusion, the 22 definitive articles were considered for the analysis of information.

3) Analysis and classification of information: Sources of information: Thanks to the Parsifal tool, the following figure was created, showing the level of contribution of the sources of information in percentages in Fig. 1.

By observing and analyzing Fig. 1, the following can be determined:

- IEEE Digital Library information source with 45% obtains a higher percentage in the graph, in this group are [5] [6] [7] [8] [9] [10] [11] [12] [13] [14].
- Science@Direct obtains 36.4% of contribution level for the present article. In this group are [15] [16] [17] [18] [19] [20] [21] [22].
- Scopus, which is another of the repositories that made the greatest contribution to the research, contributed 13.6%. This group includes [23] [24] [25].
- Finally, 4.5% belongs to Doaj with the article [26].

Years of publication: Thanks to the Parsifal tool, a graph showing the number of articles per year was generated in Fig. 2.

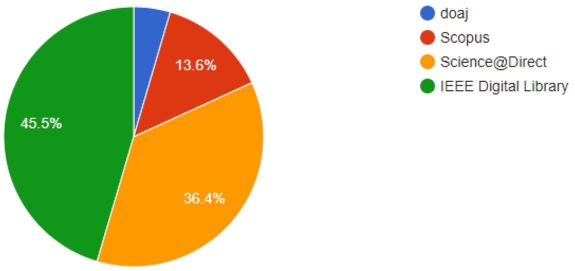


Fig. 1. Information sources.

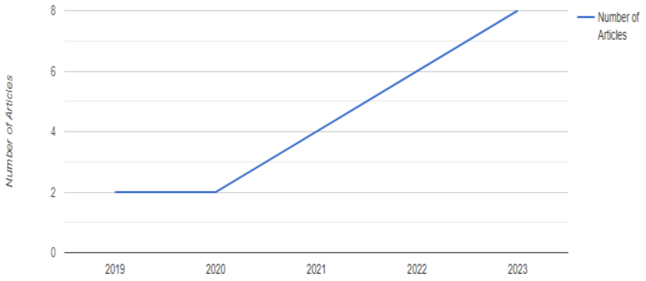


Fig. 2. Articles by year.

C. Document and Interpret the Results of the Review

The following section presents the answers to the research questions.

1) In which repositories have most of the articles on the topic of study been published?: Taking into account the information analysis and classification section, it was observed that the information source with the most articles published was the IEEE Digital Library with 10 articles, followed by Science@Direct with eight articles, these being the repositories with the greatest contribution in the field of Machine Learning in learning problems and disorders.

2) In what year were most of the articles published, and how are they interpreted?: Taking into account the section of analysis and classification of the information, it is possible to identify that the year 2023 was when more articles were published with a total of 8 articles and this was growing since 2019, this could indicate that interest in applying Machine

Learning techniques in areas such as learning disorders is growing.

3) How does Machine Learning help in the detection of learning problems and disorders?: According to the review of the articles, most of them are focused on the early detection of learning disorders, and another large percentage apply Machine Learning techniques for the classification of people with ADHD, followed by another group of articles related to the prediction of the presence of dyslexia in children. And a small percentage focus on optimizing the detection or classification of learning disorders using new and novel Machine Learning techniques, obtaining promising results.

IV. Results

The present review allowed answering the research questions, obtaining that the major source of information is the IEEE Digital Library and the year 2023 is the highest percentage of published articles indicating that more and more emphasis is being placed on the application of Machine Learning in learning problems and disorders. A brief summary of some articles will be given below:

They analyzed in [6] the classification performance of three machine learning algorithms (Naive Bayes, kNN, logistic regression) applied on the dataset of 157 children, of which 77 were ADHD patients and 80 healthy. The closest classifier k is able to predict with a high accuracy of 86% and is much better than Naive Bayes (52%) and logistic regression (66%).

This study in [15] focused on addressing data imbalance in prescreening tests for developmental dyslexia. It proposed an ensemble-based oversampling and machine learning technique to improve the detection of the minority class (dyslexic patients). The researchers used reading and writing tests, online games, magnetic resonance imaging (MRI), electroencephalography (EEG), photography, and video recording as input data. Simulation results showed that the proposed approach improved the detection accuracy of the minority class from 80.61% to 83.52%.

This research in [26] focused on identifying the neurocognitive characteristics of Attention Deficit Hyperactivity Disorder (ADHD), both in its pure presentation and in its comorbidity with other disorders. Using the Machine Learning Decision Tree (MLT) technique in the Rcran 4.2.1 software, probabilistic rules were constructed based on different neurocognitive variables. We found that children with pure ADHD showed poor performance on working memory and perceptual reasoning tasks, independent of IQ deficits. In addition, deficits in working memory were common across all ADHD presentations and comorbidities. The use of DTM allowed us to establish a clinical hierarchy and identify the most important variables in the different populations of children diagnosed with ADHD. This technique proved to be advantageous in differentiating the importance of dependent variables in the study of ADHD.

This project [5] aims to classify attention deficit hyperactivity disorder (ADHD) using machine learning techniques. The WEKA toolkit was used to perform a comparative analysis of the classification of four forms of ADHD using various machine learning algorithms. Different feature sets were experimented with, including those generated using the genetic algorithm from phenotypic data of the ADHD-200 dataset. The classification algorithms used were Logistic, Support Vector Machine (SVM), Decision Tree (DT) implemented using the J48 algorithm, Random Forest (RF), K-nearest neighbor (KNN) implemented using the instance-based learner (IBk) algorithm and multi-layer perceptron (MLP). Eight performance parameters were evaluated: accuracy, precision, recall, F-measure, Kappa statistic, root mean square error (RMSE), Mathew correlation coefficient (MCC) and area under the receiver operating characteristics curve (AUROC). The ultimate goal is to provide valuable information for the establishment of an automated diagnostic system for ADHD.

V. Discussion and Conclusion

The literature review reveals important points of convergence and divergence among the studies analyzed. First, there is a general consensus in the scientific community on the growing relevance and application of machine learning techniques, especially in the context of learning disorders and problems, such as Attention Deficit Hyperactivity Disorder (ADHD) and developmental dyslexia. All studies highlight the importance of using machine learning algorithms to improve the detection, classification and understanding of these disorders, leveraging diverse data sources, from reading and writing tests to magnetic resonance imaging and electroencephalography.

Different methodological approaches are also observed across studies. For example, while some focus on the performance of specific algorithms for ADHD classification, others address specific challenges such as data imbalance in prescreening tests for developmental dyslexia. In addition, the use of different feature sets and the application of specific techniques, such as the Machine Learning Decision Tree, highlight the diversity of strategies implemented in the research.

However, aspects related to the integration and standardization of the proposed methods remain to be addressed, as well as the validation of the results in clinical settings and the consideration of ethical factors in the development of automated diagnostic systems. The variability in the feature sets and algorithms used suggests the need to establish common protocols for the comparison and replication of results. Furthermore, the practical implementation and acceptance of these technologies in clinical and educational settings requires careful consideration of the ethical and social implications. In this sense, future research could focus on addressing these aspects to move towards the effective application of machine learning techniques in the diagnosis and treatment of learning disorders.

References

[1] A. Sans, C. Boix, R. Colomé, A. López-Sala, y A. Sanguinetti, «Trastornos del aprendizaje | Pediatría integral», 2012. Accedido: 8 de julio de 2023. [En línea]. Disponible en:

https://www.pediatriaintegral.es/publicacion-2017-01/trastornos-del-aprendizaje-2017/.

[2] J. Vaquerizo-Madrid, «Clinical assessment of attention deficit hyperactivity disorder, interview model and controversial issues», Rev. Neurol., vol. 46, n.o SUPPL. 1, 2008, doi: 10.33588/rn.46s01.2008016.

[3] S. Mariza y E. O. Alvarado, «Trastornos de aprendizaje en la educación primaria», Educación, vol. 24, n.o 2, pp. 211-215, dic. 2018, doi: 10.33539/EDUCACION.2018.V24N2.1340.

[4] M. Petticrew y H. Roberts, Systematic Reviews in the Social Sciences: A Practical Guide. Blackwell Pub, 2008. doi: 10.1002/9780470754887.

[5] J. Singh, G. Kaur, y N. Kapoor, «Classification of Attention Deficit Hyperactivity Disorder Using Machine Learning», 2022 IEEE 3rd Glob. Conf. Adv. Technol. GCAT 2022, 2022, doi: 10.1109/GCAT55367.2022.9971947.

[6] S. Saini, R. Rani, y N. Kalra, «Prediction of Attention Deficit Hyperactivity Disorder (ADHD) using machine learning Techniques based on classification of EEG signal», en 8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 782-786. doi: 10.1109/ICACCS54159.2022.9785356.

[7] K. Banumathi, G. Sudhasadasivam, B. Banurekha, N. Shahana, y K. Vaishnavi, «AI Powered Screening Aid for Dyslexic Children in Tamil», 2023 Int. Conf. Adv. Intell. Comput. Appl. AICAPS 2023, 2023, doi: 10.1109/AICAPS57044.2023.10074214.

[8] C. Sharmila, N. Shanthi, S. Santhiya, E. Saran, K. Sri Rakesh, y R. Sruthi, «An Automated System for the Early Detection of Dysgraphia using Deep Learning Algorithms», 2nd Int. Conf. Sustain. Comput. Data Commun. Syst. ICSCDS 2023 - Proc., pp. 251-257, 2023, doi: 10.1109/ICSCDS56580.2023.10105022.

[9] G. P. Pralhad, A. Joshi, M. Chhipa, S. Kumar, G. Mishra, y M. Vishwakarma, «Dyslexia Prediction Using Machine Learning», Proc. - 2021 1st IEEE Int. Conf. Artif. Intell. Mach. Vision, AIMV 2021, 2021, doi: 10.1109/AIMV53313.2021.9671004.

[10] N. Gupte, M. Patel, T. Pen, y S. Kurhade, «Early detection of ADHD and Dyslexia from EEG Signals», 2023 IEEE 8th Int. Conf. Converg. Technol. I2CT 2023, 2023, doi: 10.1109/I2CT57861.2023.10126272.

[11] M. Maniruzzaman, M. A. M. Hasan, N. Asai, y J. Shin, «Optimal Channels and Features Selection based ADHD Detection from EEG Signal using Statistical and Machine Learning Techniques», IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3264266.

[12] S. Liu et al., «Deep Spatio-Temporal Representation and Ensemble Classification for Attention Deficit/Hyperactivity Disorder», IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 1-10, 2021, doi: 10.1109/TNSRE.2020.3019063.

[13] L. Shao, D. Zhang, H. Du, y D. Fu, «Deep forest in ADHD data classification», IEEE Access, vol. 7, pp. 137913-137919, 2019, doi: 10.1109/ACCESS.2019.2941515.

[14] A. Mohd, A. M. Ali, y S. A. Halim, «Detecting ADHD Subjects Using Machine Learning Algorithm», en 2022 IEEE International Conference on Computing, ICOCO 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 299-304. doi: 10.1109/ICOCO56118.2022.10031796.

[15] S. Kaisar y A. Chowdhury, «Integrating oversampling and ensemble-based machine learning techniques for an imbalanced dataset in dyslexia screening tests», ICT Express, vol. 8, n.o 4, pp. 563-568, dic. 2022, doi: 10.1016/j.icte.2022.02.011.

[16] J. Kunhoth, S. Al Maadeed, M. Saleh, y Y. Akbari, «CNN feature and classifier fusion on novel transformed image dataset for dysgraphia diagnosis in children», Expert Syst. Appl., vol. 231, p. 120740, nov. 2023, doi: 10.1016/J.ESWA.2023.120740.

[17] N. P. Guhan Seshadri, S. Agrawal, B. Kumar Singh, B. Geethanjali, V. Mahesh, y R. B. Pachori, «EEG based classification of children with learning disabilities using shallow and deep neural network», Biomed. Signal Process. Control, vol. 82, abr. 2023, doi: 10.1016/J.BSPC.2022.104553.

[18] A. Devi y G. Kavya, «Dysgraphia disorder forecasting and classification technique using intelligent deep learning approaches», Prog. Neuro-Psychopharmacology Biol. Psychiatry, vol. 120, p. 110647, ene. 2023, doi: 10.1016/J.PNPBP.2022.110647.

[19] P. Raatikainen, J. Hautala, O. Loberg, T. Kärkkäinen, P. Leppänen, y P. Nieminen, «Detection of developmental dyslexia with machine learning using eye movement data», Array, vol. 12, p. 100087, dic. 2021, doi: 10.1016/J.ARRAY.2021.100087.

[20] Z. Mao et al., «Spatio-temporal deep learning method for ADHD fMRI classification», Inf. Sci. (Ny)., vol. 499, pp. 1-11, oct. 2019, doi: 10.1016/j.ins.2019.05.043.

[21] A. Riaz, M. Asad, E. Alonso, y G. Slabaugh, «DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI», J. Neurosci. Methods, vol. 335, p. 108506, abr. 2020, doi: 10.1016/J.JNEUMETH.2019.108506.

[22] B. TaghiBeyglou, A. Shahbazi, F. Bagheri, S. Akbarian, y M. Jahed, «Detection of ADHD cases using CNN and classical classifiers of raw EEG», Comput. Methods Programs Biomed. Updat., vol. 2, p. 100080, ene. 2022, doi: 10.1016/j.cmpbup.2022.100080.

[23] P. Drotár y M. Dobeš, «Dysgraphia detection through machine learning», Sci. Rep., vol. 10, n.o 1, pp. 1-11, dic. 2020, doi: 10.1038/S41598-020-78611-9.

[24] L. Devillaine et al., «Analysis of Graphomotor Tests with Machine Learning Algorithms for an Early and Universal Pre-Diagnosis of Dysgraphia», Sensors, vol. 21, n.o 7026, p. 7026, oct. 2021, doi: 10.3390/S21217026.

[25] M. Maniruzzaman, J. Shin, M. A. M. Hasan, y A. Yasumura, «Efficient Feature Selection and Machine Learning Based ADHD Detection Using EEG Signal», Comput. Mater. Contin., vol. 72, n.o 3, pp. 5179-5195, 2022, doi: 10.32604/cmc.2022.028339.

[26] C. Quintero-López, V. D. Gil-Vera, D. A. Landinez-Martínez, J. P. Vargas-Gaviria, y N. Gómez-Muñoz, «Predictive Neurocognitive Model of Attention Deficit Hyperactivity Disorder Diagnosis», Mediterr. J. Clin. Psychol., vol. 11, n.o 3, abr. 2023, doi: 10.13129/2282-1619/MJCP-3606.

Durian Disease Classification using Vision Transformer for Cutting-Edge Disease Control

Marizuana Mat Daud[1], Abdelrahman Abualqumssan[2], Fadilla 'Atyka Nor Rashid[3],
Mohamad Hanif Md Saad[4], Wan Mimi Diyana Wan Zaki[5], Nurhizam Safie Mohd Satar[6]

Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia[1]
Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia[2, 4, 5]
Centre for Artificial Intelligence Technology, Faculty of Information Science & Technology,
Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia[3]
Centre for Software Technology & Management, Faculty of Information Science & Technology,
Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia[6]

***Abstract*—The durian fruit holds a prominent position as a beloved fruit not only in ASEAN countries but also in European nations. Its significant potential for contributing to economic growth in the agricultural sector is undeniable. However, the prevalence of durian leaf diseases in various ASEAN countries, including Malaysia, Indonesia, the Philippines, and Thailand, presents formidable challenges. Traditionally, the identification of these leaf diseases has relied on manual visual inspection, a laborious and time-consuming process. In response to this challenge, an innovative approach is presented for the classification and recognition of durian leaf diseases, delves into cutting-edge disease control strategies using vision transformer. The diseases include the classes of leaf spot, blight sport, algal leaf spot and healthy class. Our methodology incorporates the utilization of well-established deep learning models, specifically vision transformer model, with meticulous fine-tuning of hyperparameters such as epochs, optimizers, and maximum learning rates. Notably, our research demonstrates an outstanding achievement: vision transformer attains an impressive accuracy rate of 94.12% through the hyperparameter of the Adam optimizer with a maximum learning rate of 0.001. This work not only provides a robust solution for durian disease control but also showcases the potential of advanced deep learning techniques in agricultural practices. Our work contributes to the broader field of precision agriculture and underscores the critical role of technology in securing the future of durian farming**.

Keywords—Vision transformer; durian disease; deep learning; disease control

I. Introduction

The durian fruit's popularity has surged in recent years, primarily driven by increased consumer demand, notably from China [1]. Moreover, it has found a substantial export market in Southeast Asian countries, Hong Kong, Australia, and Western nations such as United States. This upswing in the durian market can be attributed in part to the cultivation of premium varieties renowned for their exceptional flavor and consistent pulp quality. Notably, varieties like D24, D197 (Musang King), and D200 (Black Thorn) from Malaysia, as well as traditional Thai cultivars such as Monthong, Chanee, and Kanyau, have garnered significant attention and are in high demand, painting a promising future for the fruit.

Thailand maintains its position as the primary producer and exporter of durians, with other countries like Malaysia, Indonesia, Vietnam, Cambodia, and the Philippines also cultivating this unique fruit [2]. The global durian fruit trade is characterized by a dominant duopoly, with China taking the lead in imports, while Thailand leads in exports. In 2021, Thailand's durian exports reached an impressive value of 3,920 million USD, making up a significant 82.7% of the total global trade. In contrast, Malaysia's contribution ranked fourth, comprising about 0.67% of the trade volume, with a total value of 31.8 million USD. Simultaneously, China asserted its dominance in global durian imports in the same year, with an astonishing 4,240 million USD, constituting a substantial 89.4% of the overall trade. Additionally, notable participants in the market, following China's lead, included Hong Kong, Vietnam, Chinese Taipei, and Singapore, accounting for 89.4%, 5.37%, 2.43%, 0.72%, and 0.36% of the trade, respectively [3]. Fig. 1 and Fig. 2 depict the top importer and exporter of durians, respectively.

The adoption of modern agricultural practices, including drip irrigation, enhanced fertilizer formulations and application techniques, and improved cultural and postharvest methods, has significantly contributed to the increased productivity of durian farms [4]. Nevertheless, growers remain vigilant due to the persistent threat of diseases in the industry. Durian trees are susceptible to a range of diseases, such as spot cancer, base rot, base disease, seedling disease, dead tip disease, fungal infections, leaf spots, leaf blight, and fruit rot. Among these, stem rot disease, primarily caused by P. Palmivora, stands out as a particularly perilous ailment. This disease severely impairs the tree's nutrient transport system within the stem.

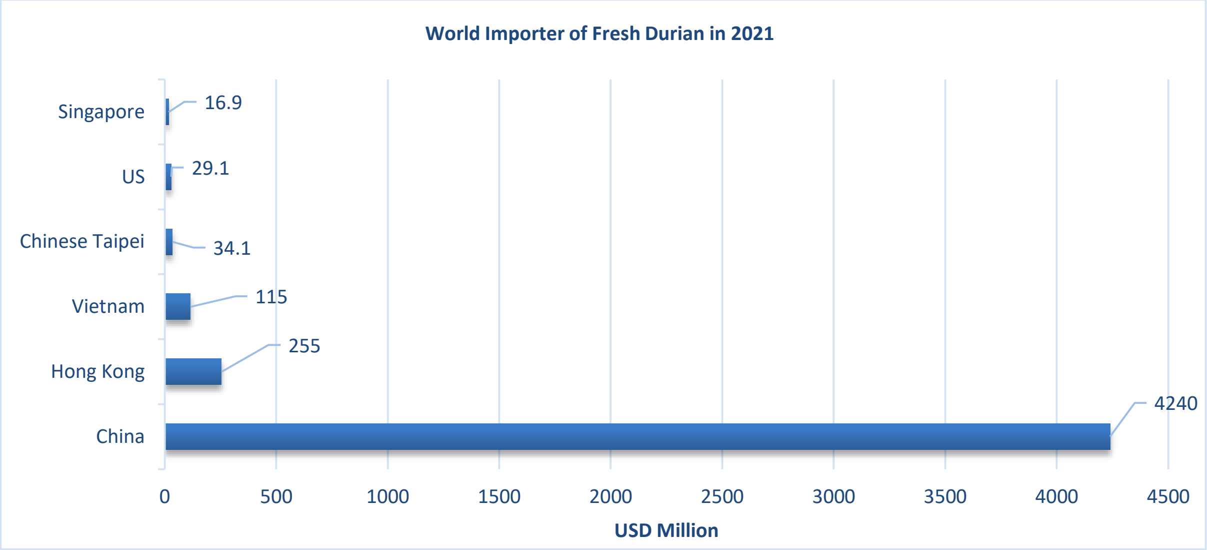


Fig. 1. World importer of fresh durian in 2021.

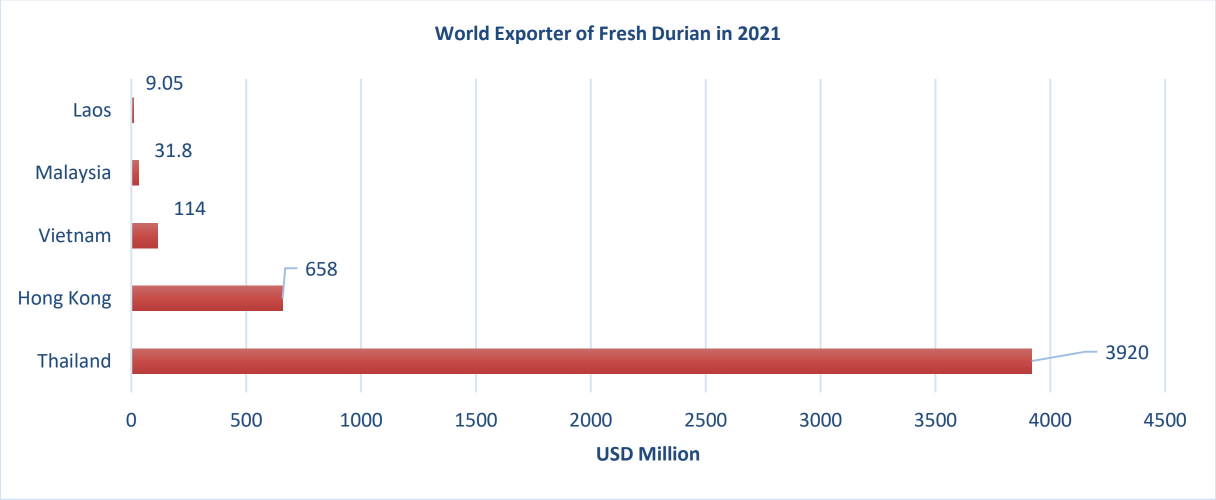


Fig. 2. World exporter of fresh durian in 2021.

Durian trees are susceptible to various diseases, such as algal diseases caused by cephaleuros virescens, characterized by the appearance of orange, rust-colored velutinous spots on the upper surfaces of leaves, twigs, and branches. Another concern is anthracnose, resulting from Colletotrichum gloeosporioides, which manifests as dark lesions on fruit and premature fruit drop. Moreover, Phomopsis leaf spot, induced by diplodia heobromae and C. Gloeosporioides, presents as necrotic, brown circular spots, approximately 1 mm in diameter, featuring dark margins and yellow halos on leaves. The sinister pink disease, attributed to erythricium salmonicolor, is marked by pinkish-white mycelial threads that envelop branches and shoots.

Additionally, postharvest fruit rots caused by Phyllosticta sp. and curvularia eragrostidis result in irregular necrotic patches in varying shades of brown. Rhizoctonia leaf blight, originating from Rhizoctonia solani, leads to water-soaked spots on leaves that coalesce to form larger, irregular, water-soaked patches, eventually drying into light brown necrotic lesions. Lastly, sooty mold and black mildew, caused by Black Mildew fungi, form a hard, lumpy crust on twigs and leaf petioles, and on fruit, they create a spongy crust on the surface. However, a particularly dangerous ailment is stem rot disease, resulting from P. Palmivora, which damages the tree's nutrient transport system in the stem.

A significant aspect of the challenges that arise in agricultural areas can be addressed by computer vision [5]. Traditionally, the detection of plant diseases heavily relied on manual inspections conducted by farmers or laborers, typically with the naked eye (Singh et al., 2017 & Petrellis, 2015). Table I presented traditional disease monitoring procedure for disease management with limitations, such as visual inspection, scouting, weather-based disease forecasting and etcetera. This method can be both laborious and repetitive, especially when dealing with tall Durian trees. However, the advent of artificial intelligence (AI) has revolutionized disease detection in various tree types, including Durian.

TABLE I. TRADITIONAL DISEASE MONITORING PROCEDURE FOR DISEASE MANAGEMENT

Disease Monitoring Procedure	Description	Limitations
Visual Inspection	Regular visual assessment of crops for symptoms of disease.	- Subject to human error and bias. - May miss early or subtle symptoms. - Time-consuming for large fields.
Scouting by Field Observers	Trained personnel systematically inspecting fields for signs of disease.	- Labor-intensive and costly. - Limited coverage and potential variations in observer expertise.
Weather-Based Disease Forecasting	Using weather data to predict disease outbreaks based on favourable conditions for pathogen development.	- Accuracy depends on the quality and availability of weather data. - Doesn't account for all factors affecting disease.
Sampling and Lab Testing	Collecting plant or soil samples for laboratory analysis to identify and confirm disease presence.	- Requires specialized equipment and expertise. - Results may not be available quickly enough for immediate action.
Disease Severity Rating Scales	Assigning numerical scores to rate disease severity, helping quantify disease progression.	- Subjective and dependent on the assessor's judgment. - Can be time-consuming, especially for large areas.
Trap Crops and Indicator Plants	Planting susceptible species near valuable crops to serve as early warning indicators of disease presence.	- May not always provide timely detection. - May require additional land and resources.
Neighbouring Farm Communication	Exchange of information among neighbouring farms about disease outbreaks or observations.	- Relies on the willingness of nearby farmers to share information. - Limited to local awareness.

In agriculture, diseases are a common occurrence across different fruit varieties. When it comes to monitoring fruit diseases, researchers and practitioners often grapple with the challenge of finding a balance between the accuracy of deep learning models and the computational resources necessary for efficient monitoring. To tackle this challenge and enhance both precision and efficiency, various deep learning architectures and techniques have been explored.

Considering there are more pixels in an image than there are words in NLP applications, the use of the attention mechanism in vision applications has been considerably more constrained due to the high computing cost [6]. This means that typical attention models cannot be applied to visuals.

II. RELATED WORKS

Transformer network applications in computer vision were recently reviewed in [7] and vision transformer (ViT) is a major step towards adopting transformer-attention models for computer vision tasks [8]. Compared to CNN-based models that consider picture pixels, using image patches as information units for training is groundbreaking. ViT uses self-attention modules to analyze the relationship between image patches included in a shared region. ViT was demonstrated to outperform CNNs in image classification accuracy given vast quantities of training data and processing resources [8].

State-of-the-art deep learning models can achieve impressive results and are well-suited for drone applications, but they come with a hefty need for computational resources during the training process. In contrast, Vision Transformer (ViT) offers a promising alternative [26]. ViT avoids using Convolutional Neural Networks (CNNs) and performs at a level similar to top-tier CNN models. ViT, a relative of the Transformer model, utilizes a smart technique called self-attention to establish a global reference for each pixel in an image during training. It breaks the image into smaller patches, assigns a position to each patch, and learns from them. In the final layers of the ViT model, the similarity between these patch representations significantly improves. Interestingly, adding more layers to the model doesn't enhance its performance [27]. Nevertheless, ViT does pose a challenge when dealing with high-resolution images due to its four-fold increase in memory requirements, making them more difficult to handle.

Numerous research studies have focused on mitigating the shortcomings of Transformer-based models which tend to fall into two main categories: hybrid models and pure Transformer enhancements. Table II illustrates both hybrid and pure transformer enhancements. Furthermore, by employing the Vision Transformer, researchers achieved a minimum of 1% higher accuracy in classifying cassava leaf diseases than well-known CNN models. They also effectively implemented this model on the Raspberry Pi 4, an edge device, showcasing the substantial potential for its application in the realm of smart agriculture [17]. To the best of our knowledge, only [19] has performed durian disease detection using deep learning approach. The durian disease classification was performed using Resnet-9 and VGG-19 where Resnet-9 was outperformed VGG-19 with accuracy of 100% and 99.11%, respectively. Recent advancements in plant disease detection have seen substantial enhancements using CNN-based models. Nevertheless, these models have limitations such as translation invariance, locality sensitivity, and a lack of comprehensive global image understanding.

TABLE II. HYBRID AND ORIGINAL VISION TRANSFORMER METHOD

Paper	ViT Techniques	Results	Limitations
[9]	Ghost-Enlightened Transformer (GeT)	98.14%	-Relies on large labelled data
[10]	PlantXViT	98.33%	-unable to maintain a lower count of Gega floating operation points.
[11]	Convolution vision Trasnformer (CvT)	87.7%	-higher accuracy will increase training and inference times and memory used.
[12]	Convolution-enhanced image Transformer (CeiT)	99.1%	
[13]	LocalVit	94.2%	
[14]	Swin Transformer	81.3%	-larger resolution needed to increase the accuracy
[15]	k-NN attention (KvT)	73.0%	-need to be paired up as the boosting agent for the vision transformer.
[16]	RegionViT	83.8%	Not stated

To overcome these challenges, this study introduces a novel approach that employs a Vision Transformer-based model for more effective plant disease classification. ViT results will be compared with ResNet-9 and VGG-19 [19] in results and discussion part. This approach combines computer vision and deep learning technologies to revolutionize agricultural production management, utilizing large-scale datasets to address current agricultural issues and improve the overall performance of agricultural automation systems, especially in Durian disease classification, thereby propelling agricultural automation equipment and systems toward a more intelligent future [18].

The paper is organized as follows: Section I presents a brief introduction to the type of durian diseases and current method used to detect the diseases. Section II delves into related works. Section III covers the methodology of ViT and how the experiment conducted. Then, Section IV presents the results and discussion of durian disease detected using ViT and Section V gives the conclusion and future work of the research.

III. Methodology

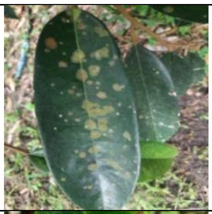
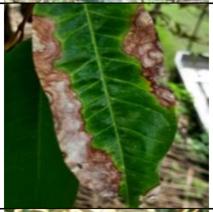
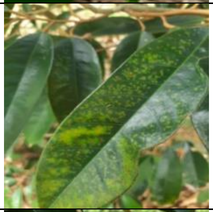
A. Dataset Preparation

In this experimental study, our primary objective was to develop and train a robust deep learning model specifically Vision Transformer capable of accurately classifying diseases that affect durian plants. Our dataset included a total of 1,344 images, which were distributed across four distinct classes. The diseases aimed to precisely classify were ‘durian_leaf_spot’, ‘durian_leaf_blight’, ‘durian_algal_leaf_spot’, and ‘durian_ healthy’, as presented in Table III [20].

To effectively manage the dataset, the original dataset is divided into two sets: training and validation. This was achieved by applying a validation split ratio of 20%, meaning that 80% of the data was designated for training purposes, while the remaining 20% was reserved for validation.

Additionally, to enhance the model's generalization and diversify the training data, data augmentation techniques was implemented. The 'ImageDataGenerator' class is provided by Keras, which facilitated various augmentations of the training data. These augmentations included random rotations of up to 40 degrees, horizontal and vertical shifts of up to 20% of the image dimensions, shearing transformations up to 20%, random zoom adjustments that could expand or contract by up to 20%, and horizontal flipping to create mirror images. When generating new pixels, the 'nearest' method was utilized. This process resulted in the creation of four augmented versions of each original image, significantly expanding the size of the training dataset.

TABLE III. Durian Disease Dataset Extracted From [20]

Dataset	Description	Total number of images	Sample of images
'durian___leaf_spot'	Images of durian leaves with leaf spot disease.	336	
'durian___leaf_blight'	Images of durian leaves with leaf blight disease	336	
'durian___algal_leaf_spot'	Images of durian leaves with algal leaf spot disease	336	
'durian___healthy'	Images of healthy durian leaves	336	

B. Durian Disease Classification using Vision Transformer (ViT)

The ViT model consists of patch creation, patch encoding, multiple Transformer layers, and a final classification head, as shown in Fig. 3.

- Patch creation: Instead of processing entire images at once, the ViT model divides each image into smaller non-overlapping patches or tiles. This patch-based approach allows the model to process large images efficiently. Each patch is treated as a separate input and processed independently by the model.
- Patch Encoding: After splitting the image into patches, each patch is encoded into a numerical representation that the model can work with. This process typically involves linearly projecting the patch's pixel values into a lower-dimensional vector, allowing the model to learn spatial relationships and features within each patch.
- Multiple Transformer Layers: The heart of the ViT model consists of multiple Transformer layers. These layers process the encoded patches and capture contextual information, enabling the model to understand how different patches relate to one another. The self-attention mechanism in the Transformer architecture is particularly crucial for this step, as it helps the model weigh the importance of different patches when making predictions.
- Final Classification Head: At the end of the ViT model, there is a classification head. This part of the model takes the information from the previous Transformer layers and makes predictions based on the features learned during the earlier stages of processing. For tasks like image classification, this is where the model assigns labels or probabilities to different classes.

The ViT model employs two optimizers, Adam and SGD (Stochastic Gradient Descent), which include a regularization technique called weight decay. Weight decay is a regularization method used to prevent overfitting in deep learning models. It works by adding a penalty term to the loss function during training, encouraging the model to have smaller weight values. Smaller weights can make the model more robust and less prone to overfitting.

To improve training efficiency, a learning rate schedule is defined. In this schedule, the learning rate, which determines how much the model's parameters are updated during training, is reduced by 50% every 10 training epochs. This gradual reduction in learning rate is a common strategy to help the model converge to a good solution without making overly large updates to its weights, which can cause instability.

During training, the ViT model periodically saves its current state as checkpoints. These checkpoints capture the model's parameters, allowing you to resume training from where you left off or use the model for inference. The saving of model checkpoints is typically based on validation accuracy, meaning that the model's performance on a separate validation dataset is used as a criterion to determine when to save a checkpoint. This ensures that the saved models are based on their ability to generalize to unseen data.

Furthermore, the training data is divided into two parts: the main training data (90%) and a validation set (10%). The main training data is used to train the ViT model, while the validation set is used to monitor the model's performance during training. This split is important for assessing how well the model is learning and for tuning hyperparameters like the learning rate. After the model has been trained, it is evaluated on a separate test dataset that the model has never seen during training. This evaluation assesses the model's performance on unseen data and provides an indication of how well it can generalize to real-world scenarios. The evaluation reports two metrics: accuracy, which measures the overall correctness of predictions, and top-5 accuracy, which indicates how often the correct label is among the top five predicted labels.

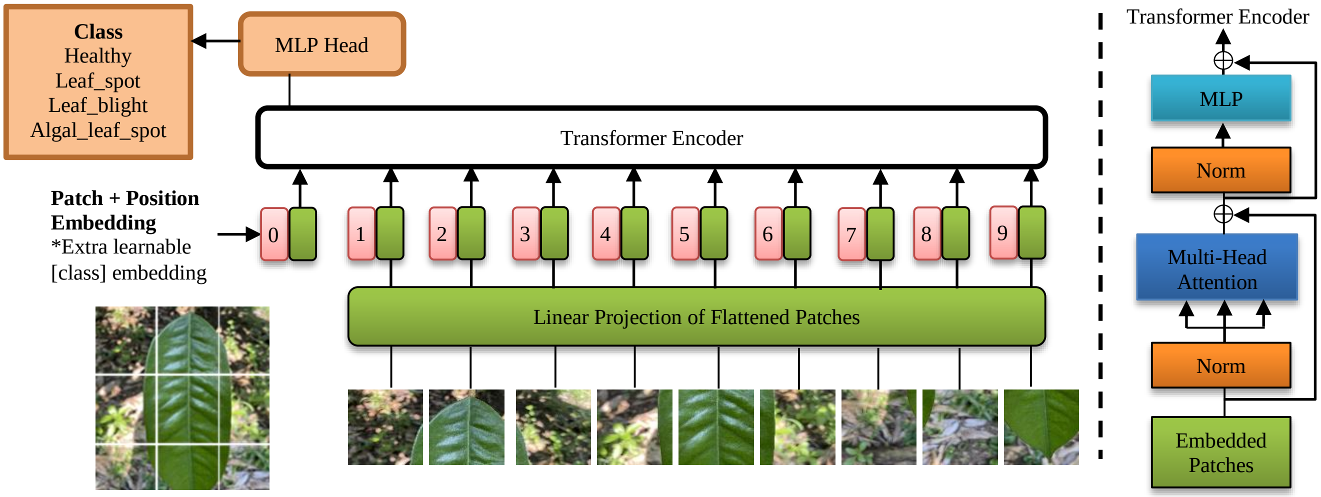


Fig. 3. Vision transformer architecture.

IV. Results and Discussion

In this study, ViT has been deployed and fine-tuned it to perform Durian Disease classification using two well-known optimization algorithms, namely, Stochastic Gradient Descent (SGD) and ADAM optimizer. The experiments encompassed a range of hyperparameters, including different learning rates (0.001, 0.005, and 0.01) and various epoch settings (20, 30, and 40). The outcomes provide valuable insights into how the model's performance varies with alterations in these key hyperparameters, shedding light on the most effective configurations for the task.

As revealed in Table IV, a validation accuracy of 94.12% was attained with the utilization of the ADAM optimizer, a maximum learning rate of 0.01, and an extended training period of 400 epochs. In contrast, Table V displays the outcomes with the SGD optimizer, where a comparable learning rate and epoch setting yielded an accuracy of 85.82%. ADAM's superior performance in this context can be attributed to its adaptability and the fusion of techniques from both momentum and RMSprop optimization. ADAM excels in scenarios involving intricate loss surfaces and fluctuating learning rates. It dynamically tailors the learning rate for each parameter, guided by historical gradient information. This adaptability often results in swifter convergence and enhanced generalization capabilities. In contrast, SGD adheres to a more conventional optimization approach. Achieving parity with ADAM's performance, particularly with complex models like ViT, often necessitates manual fine-tuning of the learning rate and other hyperparameters.

In many cases, the maximum learning rate of 0.01 might lead to faster convergence but might also make the model diverge or not settle into the optimal solution. By reducing the maximum learning rate during training (e.g., from 0.01 to 0.001), the model is allowed to fine-tune and reach a more stable and accurate solution. This phenomenon can be observed where both optimizers are performed well with maximum learning rate of 0.001 compare to 0.01.

Certainly, our training approach involved the incorporation of a learning rate schedule to foster training stability and mitigate overfitting. Simultaneously, data augmentation proved instrumental in enhancing the model's robustness by enabling it to adapt to a broader range of image conditions. Techniques such as resizing, flipping, rotation, and zooming effectively contributed to this augmentation strategy.

However, in Table VI, ResNet-9 is outperformed the other two methods, ViT and VGG-19. When evaluating why ViT might not be as good as the accuracy of ResNet-9 and VGG-19 [19], various factors come into play. First, ResNet-9 and VGG-19, as convolutional neural networks (CNNs), have been explicitly tailored for image classification tasks, boasting a proven track record in this domain. ViT, on the other hand, is a relatively newer architecture that demands substantial fine-tuning for optimal performance.

Additionally, ViT models often require more extended training schedules, specialized initialization methods, and specific architectural considerations, necessitating higher computational resources. Moreover, ViT's capacity to generalize might be challenged when confronted with smaller datasets, as its architecture is not as well-suited to such scenarios. Comparatively, ResNet-9 and VGG-19 [19] models tend to deliver robust performance with limited data, given their established history in this context.

Furthermore, the availability of pretrained weights customized for specific tasks can provide a performance advantage to ResNet-9 and VGG-19 over ViT. It's important to note that ViT is relatively more susceptible to overfitting, especially in cases involving smaller datasets or exceptionally large models. In addition to these considerations, our dataset for durian leaf disease classification is relatively small, posing a challenge for ViT's generalization capabilities. Addressing this issue would necessitate the utilization of larger models, fine-tuning with different hyperparameters, and more extensive tuning.

TABLE IV. VIT WITH ADAM OPTIMIZER

Maximum Learning Rate	Epochs	Train Loss	Validation Loss	Validation Accuracy
0.001	100	0.0647	0.2388	0.8447
0.001	200	0.0167	0.1010	0.9118
0.001	400	0.0400	0.0873	0.9412
0.005	100	0.7316	0.7451	0.6471
0.005	200	0.7406	0.7898	0.6765
0.005	400	0.7406	0.7898	0.7353
0.01	100	1.0367	1.1862	0.4706
0.01	200	1.0376	1.1377	0.4982
0.01	400	1.0270	1.0367	0.5010

TABLE V. VIT WITH SGD OPTIMIZER

Maximum Learning Rate	Epochs	Train Loss	Validation Loss	Validation Accuracy
0.001	100	0.1200	0.2689	0.8009
0.001	200	0.0951	0.2564	0.8511
0.001	400	0.0894	0.2416	0.8582
0.005	100	0.8766	0.9394	0.5843
0.005	200	0.8105	0.8324	0.6056
0.005	400	0.7791	0.8289	0.6082
0.01	100	1.1245	1.3457	0.3943
0.01	200	1.1369	1.3363	0.3973
0.01	400	1.2046	1.2298	0.4085

TABLE VI. COMPARISON OF VIT, RESNET-9 AND VGG-19 [19] USING ADAM AND SGD OPTIMIZER

Maximum Learning Rate	Adam Optimizer			SGD optimizer		
	VGG-19	**ResNet-9**	**VIT**	**VGG-19**	**ResNet-9**	**VIT**
0.001	0.8271	0.9521	0.8447	0.8542	0.8113	0.8009
0.001	0.8500	0.9797	0.9118	0.8542	0.8447	0.8511
0.001	**1.0000**	**0.9911**	**0.9412**	0.8875	0.8896	0.8582
0.005	0.8438	0.9667	0.6471	0.8708	0.8081	0.5843
0.005	0.8708	0.9792	0.6765	0.8771	0.8447	0.6056
0.005	0.8812	0.9792	0.7353	0.8708	0.8792	0.6082
0.01	0.8500	0.9729	0.4706	0.8542	0.7934	0.3943
0.01	0.8604	0.9896	0.4982	0.8542	0.8073	0.3973
0.01	0.8438	0.9896	0.5010	0.8812	0.8358	0.4085

V. CONCLUSION AND FUTURE WORKS

This research represents a significant advancement in the field of durian leaf disease detection and recognition, addressing a critical issue faced by durian farmers in ASEAN countries and beyond. The traditional manual identification of leaf diseases has been a labor-intensive and time-consuming process, posing substantial challenges to the agricultural sector's sustainability. Through the application of cutting-edge deep learning techniques and the utilization of well-established models like ViT, an automated system has been successfully developed and capable of accurately classifying and recognizing durian leaf diseases. Notably, our results demonstrate the remarkable performance, achieving an impressive accuracy rate of 94.12% when utilizing the Adam optimizer. Moreover, this research underscores the broader implications of utilizing cutting-edge machine learning techniques in agriculture. It opens the door to the development of precision agriculture systems that can revolutionize crop management practices. The implementation of ViT-based disease control not only safeguards the economic stability and food security of the Southeast Asian region but also paves the way for further advancements in the field of agriculture. With technology as a key ally, durian farmers and the agricultural sector as a whole are better equipped to overcome the challenges posed by disease and secure a more prosperous future.

ACKNOWLEDGMENT

This work was carried out with the financial support from the Ministry of Higher Education of Malaysia under the research grant LRGS/1/2019/UKM-UKM/5/2.

REFERENCES

[1] Y. Ahmad, "TECHNICAL OP-ED: Durian cultivation faces serious threat from disease caused by Phytophtora sp.," https://www.itfnet.org/v1/2018/12/management-of-phytophtora-in-durian/. (accessed on 20 Oktober 2023).

[2] T. M. UN Comtrade, "Durian Global Market Report 2018," http://www.plantationsinternational.com/ docs/durian-market.pdf. (accessed on 20 Oktober 2023).

[3] OEC, "Fruit, edible: durians, fresh," https://oec.world/en/profile/hs/fruit-edible-durians-fresh#:~:text=Exporters%20and%20Importers&text=In%202021%2C%20the%20top%20exporters,and%20Laos%20(%249.05M) (accessed on 20 Oktober 2023).

[4] 27Group, "Technology Advancement in the Durian Industry. ," https://27.group/technology-advancement-in-the-durian-industry/ (accessed on 22 Oktober 2023).

[5] M. K. Tripathi and D. D. Maktedar, "A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey," Information Processing in Agriculture, vol. 7, no. 2, pp. 183–203, 2020, doi: https://doi.org/10.1016/j.inpa.2019.07.003.

[6] R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, "Vision Transformers For Weeds and Crops Classification Of High Resolution UAV Images," CoRR, vol. abs/2109.02716, 2021, [Online]. Available: https://arxiv.org/abs/2109.02716.

[7] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," Jan. 2021, doi: 10.1145/3505244.

[8] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations, 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy.

[9] X. Lu et al., "A hybrid model of ghost-convolution enlightened transformer for effective diagnosis of grape leaf disease and pest," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 5, pp. 1755–1767, 2022, doi: https://doi.org/10.1016/j.jksuci.2022.03.006.

[10] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT." 2022.

[11] H. Wu et al., "CvT: Introducing Convolutions to Vision Transformers." 2021.

[12] K. Yuan, S. Guo, Z. Liu, A. Zhou, F. Yu, and W. Wu, "Incorporating Convolution Designs into Visual Transformers." 2021.

[13] Y. Li, K. Zhang, J. Cao, R. Timofte, and L. Van Gool, "LocalViT: Bringing Locality to Vision Transformers." 2021.

[14] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows." [Online]. Available: https://github.

[15] X. and W. F. and L. M. and C. S. and L. H. and J. R. Wang Pichao and Wang, "KVT: k-NN Attention for Boosting Vision Transformers," in Computer Vision – ECCV 2022, G. and C. M. and F. G. M. and H. T. Avidan Shai and Brostow, Ed., Cham: Springer Nature Switzerland, 2022, pp. 285–302.

[16] C.-F. Chen, R. Panda, and Q. Fan, "RegionViT: Regional-to-Local Attention for Vision Transformers." Oct. 2021.

[17] H.-T. Thai, N.-Y. Tran-Van, and K.-H. Le, "Artificial Cognition for Early Leaf Disease Detection using Vision Transformers," in 2021 International Conference on Advanced Technologies for Communications (ATC), 2021, pp. 33–38. doi: 10.1109/ATC52653.2021.9598303.

[18] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation —A review," Information Processing in Agriculture, vol. 7, no. 1, pp. 1–19, 2020, doi: https://doi.org/10.1016/j.inpa.2019.09.006.

[19] M. M. Daud, A. Abualqussan, F. A. N. Rashid and M. H. M. Saad, "Durian disease classification using transfer learning for disease management system," Journal of Information System and Technology management (JISTM), 2023.

[20] Roboflow. (2022). Durian Diseases Image Dataset. Retrieved from https://universe.roboflow.com/new-workspace-7ly0p/durian-diseases/dataset/1.

The Construction of Campus Network Public Opinion Analysis Model Based on T-GAN Model

Jianan Zhang
School of Education Science, Nantong University, Nantong 226019, China

***Abstract*—The advancement of information technology has made the internet and social media an indispensable part of modern life, but with it comes a flood of false information and rumors. The aim of this study is to develop a technology that can automatically identify campus network public opinion information, in order to protect student groups from the intrusion of erroneous information, maintain their mental health, and promote a clear campus public opinion environment. This study used the Scrapy framework to write web scraping scripts to collect campus public opinion data, and carried out cleaning and preprocessing. Then, a transformer based generative adversarial network (T-GAN) model was designed, combined with a multi-scale convolutional neural network (MCNN) structure, for public opinion analysis on campus networks. The results show that the accuracy of the dataset processed by the T-GAN model has been improved on LGBT, KNN, SVM, and RoBERTa, proving that the campus network public opinion analysis model based on the T-GAN model helps to automatically identify campus network public opinion, protect students' physical and mental health, and promote the healthy development of the campus network environment.**

Keywords—Public opinion analysis; T-GAN; feature extraction; multi-scale convolutional neural network; campus network

I. Introduction

In this era of information explosion on the Internet and social media, research has discovered a new space for people to share, communicate, and influence each other. Despite the continuous innovation of network technology, which promotes communication between people and provides unprecedented convenience in connectivity, there is also the rapid spread of unverified information hidden behind it [1]. Especially in the current era where rumors and false information are rampant, these pieces of information not only disrupt people's daily lives, but also pose a threat to social stability and order [2]. The campus environment, as a special social place, has particularly complex network public opinion management [3]. Students, as important carriers of information dissemination, frequently share insights and emotions related to numerous issues online [4]. However, the convergence of negative information on the Internet may aggravate students' rebellious mood and induce expanding foam of public opinion [5]. The constantly expanding public opinion has increased the complexity of campus management and caused interference with educational activities on campus [6]. In this context, timely monitoring and analysis of campus public opinion data is crucial for adopting effective management strategies and maintaining a harmonious and stable campus environment. In response to this issue, this study proposes a Generative Adversarial Network (GAN) model constructed using deep learning technology, aiming to conduct in-depth analysis of campus public opinion data. In addition, to improve the performance of the model in sentiment classification, this study also designed an innovative generative text data augmentation method. This method can effectively balance the distribution of data and enhance the accuracy of the model's judgment on different emotional categories. Therefore, by mining and analyzing campus public opinion data, it is helpful for schools to timely grasp the growth tendency of campus public opinion, provide important reference basis for taking effective management measures, and maintain a harmonious and stable campus environment. The novelty of this study lies in its innovative combination of Generative Adversarial Network (GAN) structure and Transformer model advantages, and the development of a deep learning model specifically designed for campus network public opinion - T-GAN. Many researchers have conducted research on this issue. Yadav A's research team has designed a deep learning model based on emotion analysis using deep learning and natural language processing techniques. Testing on popular datasets has shown significant effectiveness in solving emotion analysis problems [7]. Researchers such as Nemes L have used recurrent neural networks combined with natural language processing technology to construct an emotion classification model based on recurrent neural networks, which can classify user emotions based on keywords and demonstrate extremely high accuracy [8]. Mehta P and other scholars proposed designing an emotion mining model that utilizes machine learning and Lexicon survey methods for emotion mining. The research results show that the model can effectively match in opinion mining and sensory evaluation [9]. However, the above methods mainly address the issues of sentiment mining and sentiment classification, and research on precise sentiment classification based on campus public opinion data is relatively rare. Therefore, in response to this issue, this study innovatively utilizes deep learning technology to construct a Generate Adversarial Networks (GANs) model for in-depth analysis of campus public opinion data. Meanwhile, to improve the accuracy of the model in sentiment classification in public opinion data, a generative text data augmentation method has been studied and designed, which can expand the corpus with uneven data distribution, thereby achieving a balance in data distribution [10]. Therefore, in order to solve how to automatically identify and filter negative public opinion information in campus networks, and protect the physical and mental health of students, the establish of a campus network public opinion analysis model based on the Transformer-based Generate Adversarial Networks (T-GAN) model is studied to improve the accuracy of the model in sentiment classification and better conduct in-depth analysis of campus public opinion data. Although Yadav A's research team has successfully designed an emotion analysis deep learning model by combining deep learning and natural language processing techniques, and has demonstrated significant effectiveness in solving emotion analysis problems on popular datasets; Researchers such as Nemes L have also used a combination of recursive neural networks and natural language processing techniques to create an emotion classification model. This model classifies user emotions based on keywords and demonstrates extremely high accuracy; Mehta P and other scholars have proposed an emotion mining model that combines machine learning and dictionary research methods, and the research results show that this model is effective in opinion mining and perceptual evaluation. However, these methods mainly focus on the general issues of emotion mining and emotion classification, and there is little research on fine emotion classification based on campus public opinion data. In response to this research gap, our article aims to innovatively use deep learning techniques to establish a Generative Adversarial Networks (GANs) model, deeply analyze campus public opinion data, and explore the underlying factors that affect student emotions. Meanwhile, in order to enhance the accuracy of the model in sentiment classification of public opinion data, this study also designed a generative text data augmentation method. This method achieves data distribution balance by expanding the corpus with uneven data distribution. Through this approach, we hope that the model can more accurately understand and classify emotional tendencies in campus public opinion, provide an effective emotional analysis tool for campus managers, help them identify and respond to emotional dynamics in student groups in a timely manner, and thus maintain a healthy and harmonious campus environment. Overall, the main objective of this article is to construct a generative adversarial network model for analyzing campus public opinion data using deep learning techniques. Design and implement a text data augmentation strategy to balance data distribution and improve the accuracy of sentiment classification. Verify the effectiveness of the model on campus public opinion data and explore its potential positive impact on the campus environment. The overall structure of the entire article is divided into the following parts: Section I is the introduction, which outlines the background, importance, and purpose of the research. Emphasis was placed on the impact of the information age on the mental health and value development of students, as well as the necessity of automatically identifying campus public opinion information. Section II gives details about the design of a T-GAN based campus network. Section III delves into the verification of a campus network and Section IV concludes the paper.

II. Design of a T-GAN-Based Campus Network Public Opinion Analysis Model

The GAN model is the foundation of subsequent research in this paper. This chapter focuses on the collection and enhancement of campus network public opinion data based on the T-GAN model, and embeds a mixed MCNN structure in the network model for public opinion data analysis. The model is validated through experiments.

A. Collection and Enhancement of Campus Network Public Opinion Data Based on T-GAN Model

In the methodology section of this study, the reason for choosing to use a combination of Generative Adversarial Network (GAN) structure and Transformer model is multifaceted. Firstly, this combination fully utilizes the ability of GAN to generate authentic data and the powerful performance of Transformer in processing sequential data, making it particularly suitable for processing text data. Secondly, the current focus of research is on automatically identifying public opinion emotions in campus networks, which typically involve data in the form of text [11]. The Transformer model excels in understanding text sequences. In addition, campus public opinion data often suffers from class imbalance, and the generation ability of GANs can effectively enhance the sample of niche categories to provide richer learning materials, thereby improving the generalization ability and accuracy of the model. The participants are a group of campus network users, especially students in the cognitive development stage. Student groups are usually active in the online space, expressing opinions and feelings related to their personal life and learning experiences, which reflect their true emotions and may have an impact on others. Therefore, collecting data from it is representative and rich in valuable information such as user emotions and attitudes.

The data type to be collected in the study will mainly be public opinion information in text format, such as posts, comments, Weibo, etc. The characteristics of the required data collection include: 1. Text content: specific user statements and comments. 2. Timestamp: The time information of a speech, used to analyze public opinion dynamics. 3. User information: Basic information of the speaker, such as registration time, historical speeches, etc. (subject to privacy and regulatory restrictions).

In order to complete data collection and processing, the following tools will be used in the study: 1. Crawler tools, such as Scrapy, are used to automatically crawl target text data from online platforms and social media. 2. Data preprocessing tools: including text cleaning, denoising, word segmentation, and encoding, to prepare data formats suitable for model input. 3. Deep learning frameworks, such as TensorFlow or PyTorch, are used to construct and train T-GAN and MCNN models. 4. Evaluation and analysis tools: such as Sklearn, used for evaluating model performance and statistical analysis of results. Through this comprehensive method and corresponding tools, this study will be able to accurately capture and analyze emotional public opinion in campus networks, provide a basis for subsequent intervention measures, and thus maintain a healthy online environment.

Currently, the most commonly used social comment platform for students is Weibo, and many schools also set up official accounts on Weibo to post school related information [12]. Therefore, this study will focus on collecting public opinion data from Weibo platforms. Due to the complexity of online comment data, efficient data collection methods are needed. Traditional manual data collection methods are inefficient, so this study utilizes crawler technology to quickly collect data [13]. Through crawler technology, text data can be automatically crawled from Weibo platforms. The specific process of crawling text data is expressed in Fig. 1.

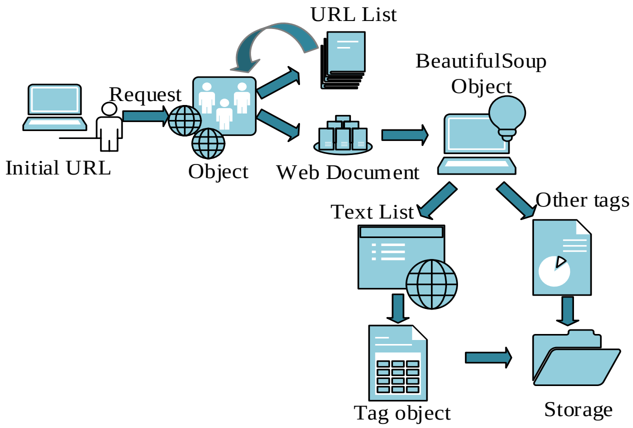


Fig. 1. The specific process of crawling text data.

From Fig. 1, when using crawler technology to crawl campus comment data, the first step is to initialize the Uniform Resource Locator (URL) to send the request. This study uses the Scrapy framework to write crawler scripts, starting from Weibo websites, and automatically accessing URL links after splitting to obtain Hyper Text Markup Language (HTML) resources in comment pages. Next, the Xpath parser is utilzied to quickly locate the target node in the web page text data, and save the HTML text attributes that need to be parsed as BeautifulSoup objects through regularization matching. This step helps to parse and manipulate complex HTML documents, saving the crawled text information in Excel or Csv format for subsequent data analysis [14]. After obtaining the crawled text information, it is necessary to clean and preprocess it for subsequent manual annotation and production of test datasets.

The crawled Weibo comment text data contains meaningless text such as emoticons, images, and punctuation marks, which need to be cleaned and deleted to improve the quality of the text data [15]. Regular matching method is an efficient, convenient, powerful, and flexible matching algorithm with good cross platform performance. Therefore, this study uses regular matching method to batch delete and preprocess the remaining text information, including removing stop words and segmenting Chinese text. During word segmentation, new words that ware not included in the vocabulary may be encountered. Therefore, this study collects 183 new words from online platforms and enters them into the vocabulary library. In addition, this study also utilizes a hidden Markov model to analyze the possible idioms composed of Chinese characters, and calculates the paths through Viterbi. Comment texts on online platforms often appear in the form of short dialogues, which contain non-standard words such as oral language and internet buzzwords. Therefore, it needs to pay attention to these characteristics during the word segmentation process [16]. By analyzing the possible idioms composed of Chinese characters using hidden Markov models, new words and idioms in comment texts can be more accurately identified, improving the accuracy and precision of text segmentation. Preprocessed text data can be transformed into data that computers can recognize and operate on using text representation technology. The Word2Vec model is an associated model that can generate word vectors and can be utilized for training to reconstruct linguistic word texts. By training the Word2Vec model, words can be mapped into vector space for computation and data analysis. The structure of the Word2Vec model is indicated in Fig. 2.

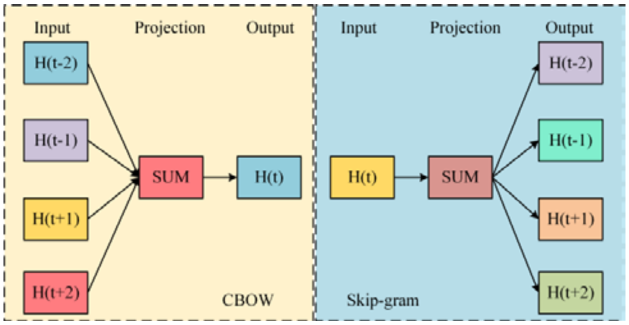


Fig. 2. The structure of the Word2Vec model.

In Fig. 2, the structure of the Word2Vec model includes two parts: Skip gram and CBOW. The Skip gram model can predict the $H(t-1), H(t-2), H(t+1), H(t+2)$ of the preceding and following text with the known head word $H(t)$. This model's goal is to infer the current word through contextual information [17]. In contrast, the CBOW model adopts the opposite approa ch, which can infer the central word from known contextual words. The training of CBOW model can be seen as a language model task, predicting the central word through a large amount of text corpus to preserve important semantic information of the text. The Skip gram and CBOW models together form a bidirectional language model that can help better understand and process natural language. The objective function expression of Skip gram is shown in Eq. (1).

$$O_{Skip} = \frac{1}{T}\sum_{t=1}^{dk}\sum_{-c\le j\le c, j\ne 0}\log P_{Skip}(H_i|H_{i+1}) \tag{1}$$

In Eq. (1), O_{skip} represents the objective function of Skip gram, and T represents time, c means the context window's size during model training. d_k represents the dictionary's size. H_i represents the word vector. $P_{Skip}(H_i|H_{i+1})$ represents the path probability of the word vector in the Skip gram model. The expression is shown in Eq. (2).

$$P_{Skip}(H_i|H_{i+1}) = \frac{\exp(<v_i, v_{i+j}>)}{\sum_{k=1}^{dk}\exp(<v_k, v_i>)} \tag{2}$$

In Eq. (2), $v_{i,}v_k$ represent the output and input vectors, respectively. The objective function expression of the CBOW model is shown in Eq. (3).

$$O_{CBOW} = \frac{1}{T}\sum_{t=1}^{dk}\sum_{-c\le j\le c, j\ne 0}\log P_{CBOW}\left(H_i|H_{i+j}\right) \tag{3}$$

In Eq. (3), O_{CBOW} represents the objective function of the CBOW model, $P_{CBOW}(H_i|H_{i+1})$ represents the path probability of the word vector in the CBOW model, as shown in Eq. (4)

$$P_{CBOW}\left(H_i\middle|H_{i+j}\right)=\frac{\exp\left(<v_i,v_{i+1}>\right)}{\sum_{k=1}^{dk}\exp\left(<v_k,v_{i+j}>\right)} \tag{4}$$

The word vector generated by the Word2Vec model is static, meaning that for the same word, its vector representation remains the same regardless of the context. In contrast, the word vectors generated by the BERT model are dynamic, meaning that the vector representation of the same word may change in different contexts [18]. This is because that the BERT model utilizes a large number of self-attention mechanisms, which can adaptively focus on different words in different contexts, thereby generating different word vectors. In addition, the BERT model differs from traditional temporal networks in that it does not emphasize the order relationship between words. In the BERT model, the vector representation of each word is independent of other words, making it more flexible to handle text data. The mathematical expression of the BERT model is denoted in Eq. (5).

$$\begin{cases} E_{(pos,2i)}=\sin\left(\dfrac{pos}{1000^{\frac{2i}{emb_{\dim}}}}\right) \\ E_{(pos,2i+1)}=\cos\left(\dfrac{pos}{1000^{\frac{2i+1}{emb_{\dim}}}}\right) \end{cases} \tag{5}$$

In Eq. (5), $E_{(pos,2i)}$ and $E_{(pos,2i+1)}$ represent odd and even position matrices, respectively, and $emb_{\dim}$ represents the embedding dimension [19].

The technology used in this article - the T-GAN model that combines Generative Adversarial Networks (GAN) and Transformer models - has the following advantages: 1) Data augmentation ability: GAN's unique data generation ability can expand and balance the dataset by generating realistic text data, which is particularly beneficial for handling imbalanced data distribution and insufficient training data. 2) High text processing performance: The Transformer model has excellent sequential data processing capabilities, making it particularly suitable for analyzing text data. Its self-attention mechanism can capture long-distance dependencies and enhance the model's understanding of context. 3) High model accuracy: Compared to traditional deep learning and machine learning models, T-GAN demonstrates higher accuracy and recall in text sentiment classification tasks, indicating superior classification performance of the model. 4) Strong generalization ability: Due to the combination of two powerful neural network models and extensive testing on multiple datasets, T-GAN exhibits good generalization ability, which means it can adapt to different types of text data analysis tasks. 5) Real time analysis capabilities: T-GAN is suitable for real-time data analysis, helping managers monitor and respond to emotional fluctuations in the network in a timely manner. It is particularly valuable for environments that require quick response, such as campuses. 6) Improving user experience: Automated sentiment classification tools can reduce the burden of manual review, improve information supervision efficiency, and thus improve the user experience on campus network platforms. 7) Contribution to social management: This technology can analyze and identify emotional tendencies in a large amount of online public opinion data, providing strong technical support for social public opinion analysis and crisis prevention.

B. Public Opinion Data Analysis Based on Hybrid Embedded MCNN Structure

GAN is a powerful generation model that can generate new sentences related to a given corpus topic [20]. This newly generated sentence is closer to real data, while ensuring data security while also preserving the smoothness of the original data as much as possible. This characteristic enables GAN to effectively improve the classification performance of the network. However, traditional GAN models mainly rely on the LSTM layer as a generator in structure, which has some limitations, such as high computational costs and a lack of interactivity between information. To address these limitations, a self-attention-based autoencoder Transformer is studied to replace the LSTM layer. Transformer, due to its ability to learn long-term dependencies, can effectively address the aforementioned limitations. The structure of the T-GAN model is expressed in Fig. 3.

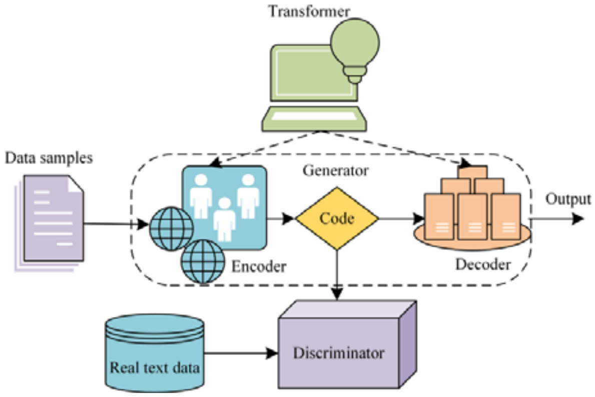


Fig. 3. The structure of T-GAN model.

In Fig. 3, the structure of the T-GAN model includes two main parts: a generator and a discriminator [21]. The generator can randomly generate vectors from space as input, generating pseudo samples with the same dimensions as real samples. The discriminator can use real and pseudo samples for training and predict a binary value to determine whether the input sample is real [22]. The T-GAN model adopts an alternating optimization method during the training process, which trains the generator network and discriminator network alternately to solve the maximum minimum game problem until both networks reach a convergence state. The mathematical expression of the maximum minimum game is expressed in Eq. (6).

$$\underset{G}{Min}\,\underset{D}{Max}\,V(G,D)=F_{x\sim Pdate(x)}\left[\log D(x)\right]+F_{z\sim noise(x)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \tag{6}$$

In Eq. (6), G, D denote the generator and discriminator. x represents the real sample. $D(x)$ expresses the real data. z means the noise vector sample. $G(z)$ refers to the generated sample. $F_{x\sim Pdate(x)}$ indicates the discriminator that follows the real data, and $F_{z\sim noise(x)}$ represents the generator and discriminator that follows the random data [23]. The Transformer architecture is embedded with a VAE module, which consists of two sub modules: encoder and decoder. The mathematical expression of the encoder is denoted in Eq. (7).

$$z \sim Enc(x) = q(z|x) \tag{7}$$

In Eq. (7), $q(z|x)$ represents the encoding space, and $z \sim Enc(x)$ represents encoding the sample x into the sample z. The mathematical expression of the decoder is expressed in Eq. (8).

$$x \sim Dec(z) = p(x|z) \tag{8}$$

In Eq. (8), $p(x|z)$ represents the decoding space, and $x \sim Dec(z)$ represents the decoding of sample z to sample x. The attention mechanism in the Transformer architecture identifies and highlights key values by adjusting weight values, and the process of finding key values is denoted in Eq. (9).

$$O = Attention(Q, K, V) \tag{9}$$

In Eq. (9), O represents the output of the attention mechanism. Q represents the query vector. V represents the key value. K represents the key value. Due to the use of multi head attention mechanism in the Transformer architecture, the three vectors Q, K, V will be projected into different subspaces, as shown in Eq. (10).

$$MH(Q, K, V) = concat(H_1, H_2, ..., H_n)W^0 \tag{10}$$

In Eq. (10), MH represents the multi head attention mechanism. H represents the subspace. n represents the degree. W^0 represents the initial matrix. By calculating the self-attention function in different subspaces, the projection output of each subspace can be obtained. Then, these output results are concatenated and the final output is obtained through projection, as shown in Eq. (11).

$$O_{MH} = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{11}$$

In Eq. (11), O_{MH} represents the projection output processed by the multi head attention mechanism. W_i^Q, W_i^K, W_i^V represents the linear projection of the corresponding response Q, K, V vector. The generator structure in the T-GAN model enhances the language model by introducing an LSTM layer to improve the attention mechanism of the decoder [24]. The generator structure of the T-GAN model is shown in Fig. 4.

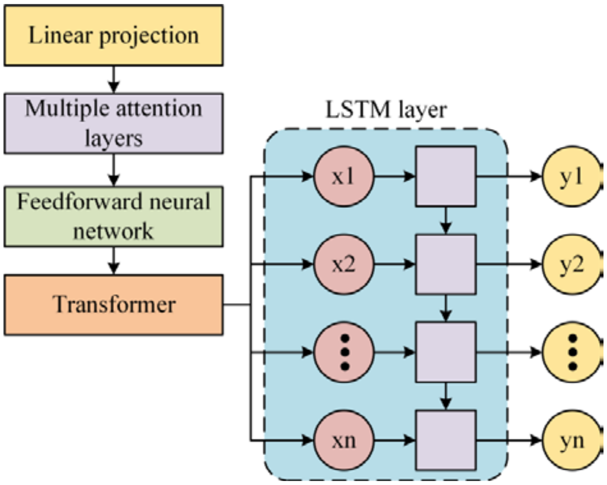


Fig. 4. The generator structure of T-GAN model.

As shown in Fig. 4, in the generator structure of the T-GAN model, the output results generated by the Transformer decoder are input into the LSTM layer for language model enhancement. This design enables the model to better focus on more advanced semantic and stylistic features, thereby generating more realistic and natural samples. The data text enhanced by the T-GAN model requires the use of MCNN for feature extraction [25]. In this process, the data text is mixed and embedded as input to add more forms of representation from the feature extraction level, thereby improving model performance. The structure of mixed embedded MCNN is shown in Fig. 5.

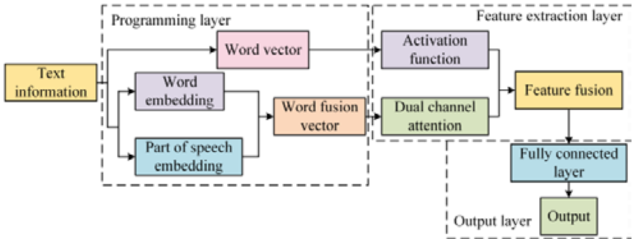


Fig. 5. The structure of mixed embedded MCNN.

As shown in Fig. 5, the structure of hybrid embedded MCNN consists of three parts: encoding layer, feature extraction layer, and output layer [26]. In the encoding layer, multi-channel input is used to improve the representation ability of short dialogue text [27]. And through part of speech embedding and word embedding, the association of parts of speech in the text is strengthened to generate mixed word vectors. The expression of the mixed word vector is shown in Eq. (12).

$$\begin{cases} w_{pi} = w_i \oplus pos_i \\ W_p = [w_{p1}, w_{p2}, ..., w_{pm}]^T \end{cases} \tag{12}$$

In Eq. (12), w_{pi} represents the mixed word vector. W_p represents the mixed word vector channel. m represents the embedding sequence [28]. The feature extraction layer performs feature extraction in the form of dual channels, as shown in Eq. (13).

$$\begin{cases} Vc = Conv(c) \\ vc = GMP(Vc) \end{cases} \tag{13}$$

In Eq. (13), Vc represents the feature vector obtained after convolution operation. c represents the word vector. GMP represents global maximum pooling. vc represents the feature vector obtained after global maximum pooling operation. The feature vectors of words and mixed words obtained through two channels are shown in Eq. (14).

$$\begin{cases} V_\alpha = \sum_{i=1}^{m} \alpha c_i \times Vc \\ V_\beta = \sum_{i=1}^{m} \beta c_i \times Vc \end{cases} \tag{14}$$

In Eq. (14), V_α, V_β represents the feature vectors of the word and the feature vectors of the mixed word, respectively. These two feature vectors are concatenated in the output layer, then input into the fully connected layer for fusion, and then processed using a Softmax classifier to obtain the classification probability as shown in Eq. (15).

$$P = soft\max(\omega\lfloor V_\alpha \oplus V_\beta \rfloor + \mu) \tag{15}$$

In Eq. (15), P represents the classification probability of the output. ω represents the weight transformation matrix. $\oplus$ represents the connection operation. μ represents the bias vector [29].

III. Verification of Campus Network Public Opinion Analysis Model Based on T-GAN Model

To experimentally validate the T-GAN-based campus network public opinion analysis model, this chapter first set the experimental environment and parameters, and analyzed the impact of experimental parameters on the performance of the model [30]. Finally, the performance of the T-GAN model was verified.

A. Experimental Environment and Parameter Settings

All experiments in this article were conducted in the Python 3.6.7 environment. The dataset used was Senti Large, with 20% of the data samples used as the test set and the other 80% used as the training set. The Senti Large dataset is an enhanced dataset generated through T-GAN based on publicly available Sentiment datasets. The Sentiment dataset is mainly used for sentiment analysis, which is a natural language processing technique aimed at identifying and extracting subjective information from text. This type of information may include the author's emotions, emotions, or opinions on a specific topic or product, which can be positive, negative, or neutral. The Sentiment dataset mainly focuses on text data, as the core of sentiment analysis is to understand and analyze emotions or emotions in the text. Therefore, this type of dataset does not contain image data. The main application scenarios of this dataset include product review analysis, social media sentiment tracking, and public opinion research. Overall, the Senti Large dataset is an enhanced text dataset for sentiment analysis, utilizing advanced T-GAN technology to expand the original Sentiment dataset to improve model performance and generalization ability. To evidence the generalization ability of the model, the study also added a manually calibrated Weibo data sample set as a pure test set. The careful setting of experimental parameters will directly affect the training effect of the model and the final classification results. To optimize the effectiveness of the model, the following parameters were set in the experiment based on the characteristics of the T-GAN model. The experimental environment configuration and parameter settings are indicated in Table I. As shown in Table 1, in order to conduct this experiment, corresponding parameters need to be configured in a specific environment. Firstly, the experiment should be conducted on the Windows 10 operating system to ensure sufficient memory capacity. Here, 64GB was chosen as it can effectively handle large datasets and run deep learning models. The graphics processor used in the experiment is NVIDIA's TITAN BLACK GPU, which can accelerate the training and prediction of deep learning models. Meanwhile, the study chose Intel Core i5-4460 as the central processing unit. PyTorch 1.8.1 was chosen as the deep learning framework in terms of software, and the CUDA11.1 platform and API were utilized to fully utilize NVIDIA GPU for computation. In terms of parameter settings, research has set vector spaces with dimensions of 30 and 300 for part of speech vectors and word vectors, respectively. The learning rate is set to 0.002, the number of convolutional kernels is 256, and the batch size is 16. In order to prevent overfitting of the model, the study set the Dropout ratio to 0.5, which means that half of the neurons will be randomly ignored during the training process. These configurations will ensure the smooth progress of the experiment and the accuracy of the results.

TABLE I. EXPERIMENTAL ENVIRONMENT CONFIGURATION AND PARAMETER SETTINGS

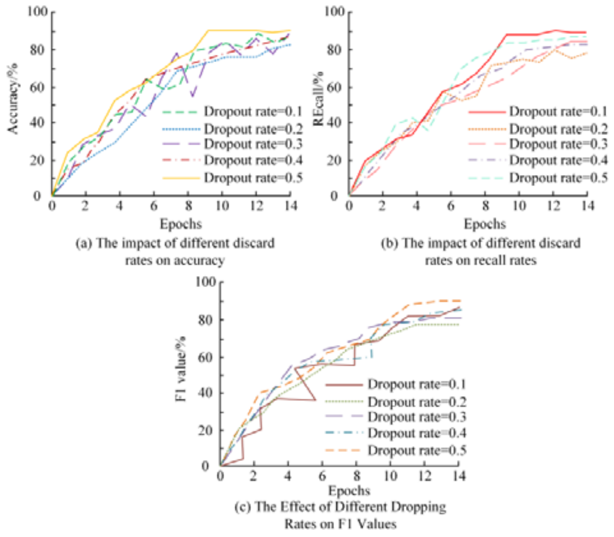
Experimental environment	Configuration	Parameter Description	Paramete r value
OS system	Windows10	Part of speech vector dimension	30
Memory	64GB	Word Vector Dimension	300
GPU	NVIDIA TITAN BLACK GPU	Learning rate	0.002
CPU	Intel(R)Core(TM)i 5-4460	Number of convolutional kernels	256
PyTorch framework	PyTorch 1.8.1	Batch size	16
CUDA Framework	CUDA11.1	Dropout rate	0.5

B. The Effect of Experimental Parameters on Model Performance

To delve deeper into the impact of parameter dropout rate on model prediction results, the experiment set the dropout rate range between $[0.1,0.2,0.3,0.4,0.5]$. After training, the impact of different dropout rates on prediction performance is shown in Fig. 6. From Fig. 6 (a), when the discard rates were 0.5 and 0.2, the accuracy reached the highest of 90.1% and the lowest of 81.8%, respectively. From Fig. 6 (b), when the discard rates were 0.1 and 0.2, the recall rates reached the highest of 89.9% and the lowest of 78.8%, respectively. From Fig. 6 (c), when the discard rates were 0.5 and 0.2, the F-value reached the highest of 90% and the lowest of 78.3%, respectively. Overall, when the discard rate was 0.5, all performance indicators of the model reached the optimal state.

The batch size is the amount of data input into the model at once when training the corpus, and the setting of this parameter has a huge impact on the training efficiency of the model. If the batch size is set too large, overfitting may occur, increasing training costs, while if the batch size is set too small, it may reduce training efficiency. Therefore, choosing an appropriate batch size is crucial. The range of batch size parameters selected for the experiment was $[4,8,16,32]$, and the impact of batch size on accuracy obtained through training is shown in Fig. 7. From Fig. 7, when the batch size was 16, the accuracy of the model reached its highest value, at 90%. In contrast, when the batch size was 4, 8, and 32, the accuracy of the model was 84.8%, 89.1%, and 82.3%, respectively, which increased by 5.2%, 0.9%, and 7.7% compared to the accuracy of the batch size of 16, respectively. Overall, the most suitable value for the batch size parameter was 16. This setting could effectively improve the accuracy of the model while ensuring its training efficiency.

In the learning process of neural networks, parameter learning rate plays a crucial role. The learning rate can help update the weight of the model through backpropagation, thereby achieving reasonable adjustment of the weight. To deeply explore the influence of learning rate on the loss function, the experiment plotted the influence of parameter learning rate on the loss function as shown in Fig. 8. From Fig. 8, when the learning rate was less than 0.002, the value of the loss function showed a significant downward trend as the learning rate increased. When the learning rate was 0.002, the decline rate of the loss function was the fastest. This observation indicated that when the learning rate was 0.002, the model could more effectively optimize weights, thereby achieving a better level of loss function.



(a) The impact of different discard rates on accuracy

(b) The impact of different discard rates on recall rates

(c) The Effect of Different Dropping Rates on F1 Values

Fig. 6. The impact of different discard rates on prediction performance.

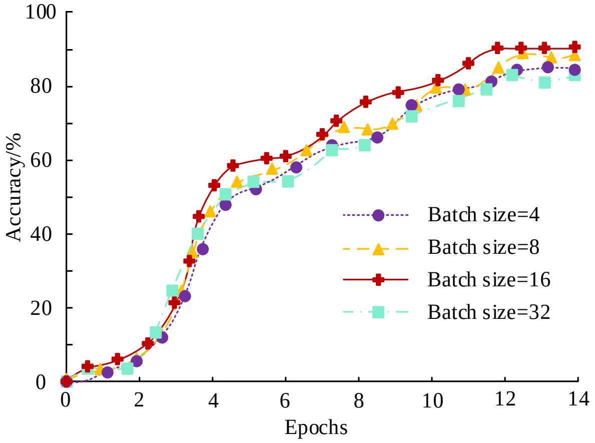


Fig. 7. The impact of parameter batch size on accuracy.

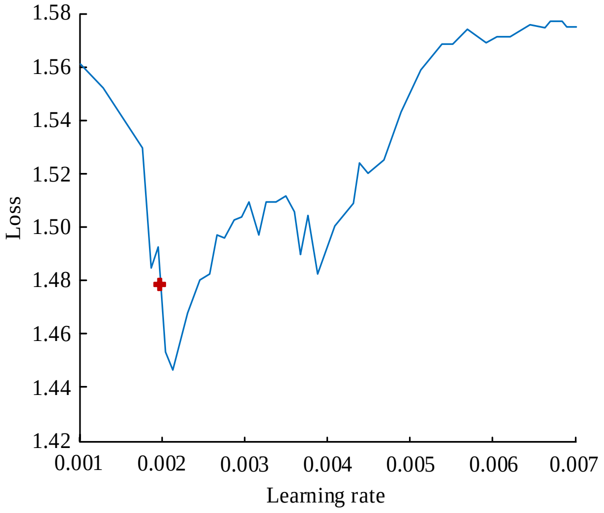


Fig. 8. The influence of parameter learning rate on the loss function.

C. Performance Verification of T-GAN Model

To visually demonstrate the performance of the T-GAN model, the new data generated by the T-GAN model was projected onto the same two-dimensional plane as the real data. The data projection is shown in Fig. 9. From Fig. 9, the new data generated by the model highly overlapped with the real data on the projection plane, which strongly proved the accuracy of the model's prediction. This phenomenon indicated that the model could capture the main features of the data and generate results that are highly similar to real data.

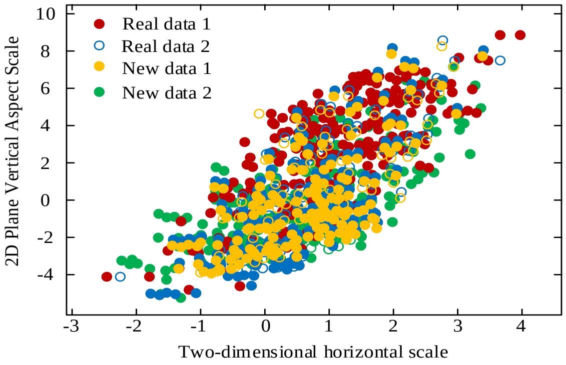
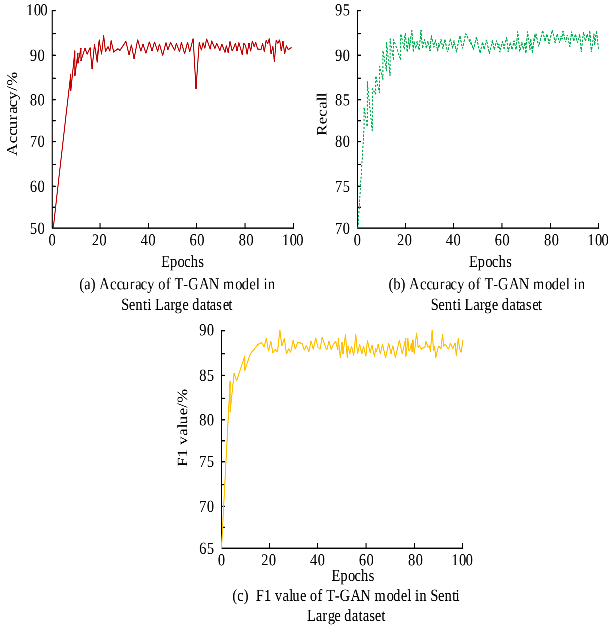


Fig. 9. Data projection map.

To prove the effectiveness of the T-GAN model, it was trained on the Senti -Large dataset and its predictive performance was evaluated. The predictive performance results of the T-GAN model are shown in Fig. 10. From Fig. 10, the T-GAN model converged rapidly on the training set, approached convergence after nearly 20 iterations, and achieved excellent predictive performance. Specifically, the T-GAN model achieved an accuracy of 92%, a recall rate of 91.8%, and an F1 value of 88.7% on the training set. These indicators all indicated that the T-GAN model performed well in predicting performance on the training set.



(a) Accuracy of T-GAN model in Senti Large dataset

(b) Accuracy of T-GAN model in Senti Large dataset

(c) F1 value of T-GAN model in Senti Large dataset

Fig. 10. Prediction performance results of T-GAN model.

In order to further verify the superior performance of the T-GAN model, state-of-the-art text feature extraction models such as LGBM, KNN, SVM, and RoBERTA [20] were used to conduct comparative experiments on the test set and training set [31]. The accuracy comparison of different models is shown in Table II. From Table II, in the training set, the accuracy of the T-GAN model was 91.56%, which was 8.89%, 6.78%, 15.87%, and 1.18% higher than models such as LGBM, KNN, SVM, and RoBERTA, respectively. In the test set, the accuracy of the T-GAN model was 92.89%, which increased by 8.93%, 8.25%, 19.01%, and 1.57% compared to models such as LGBM, KNN, SVM, and RoBERTA, respectively.

TABLE II. COMPARISON OF ACCURACY BETWEEN DIFFERENT MODELS

Model	Test set accuracy/%	Training set accuracy/%
T-GAN	92.89	91.56
LGBM	83.96	82.67
KNN	84.64	84.78
SVM	73.88	75.69
RoBERTa	91.32	90.38

In order to verify the effectiveness of the T-GAN model in physical argumentation, a series of application examples were conducted to verify its effectiveness. Taking campus cafeteria public opinion as an example, when food safety issues trigger public opinion, public opinion data mainly comes from social platforms such as Weibo, WeChat, and campus forums. The study utilized the T-GAN model to conduct in-depth public opinion analysis on the text data of these platforms, and successfully extracted internet buzzwords closely related to campus cafeteria public opinion. The extraction of campus cafeteria public opinion text using the T-GAN model is shown in Fig. 11. From Fig. 11, it can be seen that the model has successfully extracted five key network buzzwords: food safety, hygiene issues, poisoning incidents, inadequate supervision, and school responses. The frequency of these hot words appearing on social media platforms is 8054, 12567, 11067, 9104, and 11972, respectively. This data fully demonstrates that the T-GAN model performs well in extracting text information closely related to campus cafeteria public opinion. Through comprehensive analysis, it can be found that the text content extracted by the T-GAN model is highly consistent with the actual campus cafeteria public opinion. This further validates the superior performance of the T-GAN model in practical applications, proving that it can effectively extract key information closely related to food safety public opinion from massive text data.

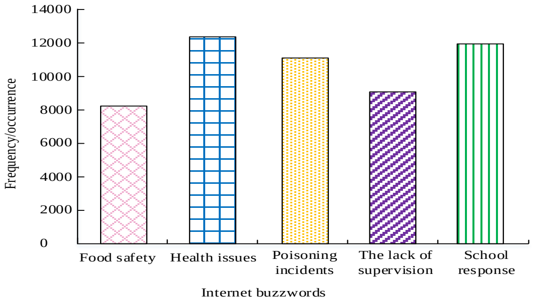


Fig. 11. Extracting public opinion text from campus canteen using T-GAN model.

IV. Conclusion

As the speed growth of network technology, the problem of spreading rumors and false information on social media is becoming increasingly prominent. To effectively and automatically identify and filter negative public opinion information in campus networks to protect students' physical and mental health, this study innovatively constructed a T-GAN model by combining the advantages of GAN network structure and Transformer model. At the same time, a hybrid embedding method using MCNN was used to conduct in-depth research on public opinion data as the analysis object. The results showed that when the discard rate was 0.5, the accuracy of the model reached the highest level of 90.1%, and the recall rate also reached the highest level of 89.9%. In addition, when the batch size was 16, the accuracy of the model also reached its highest value, at 90%. This indicated that the performance of the model was relatively stable for different discard rates and batch sizes. In the Senti-Large dataset, the accuracy of the T-GAN model reached 92%, the recall rate was 91.8%, and the F1 value also reached 88.7%. In the training set, the accuracy of the T-GAN model was 91.56%, which improved by 8.89%, 6.78%, 15.87%, and 1.18% compared to models such as LGBM, KNN, SVM, and RoBERTA, respectively. In the test set, the accuracy of the T-GAN model was 92.89%, which increased by 8.93%, 8.25%, 19.01%, and 1.57% compared to models such as LGBM, KNN, SVM, and RoBERTA, respectively. In summary, the T-GAN model performs best when the discard rate is 0.5 and the batch size is 16. Meanwhile, compared to other models, the T-GAN model has shown certain advantages in accuracy. These results indicate that the T-GAN model studied has good generalization performance and adaptability, and is worthy of further research and application. However, the research only analyzed the Chinese text, and the results are not comprehensive enough. This aspect still needs further improvement. Given the improvement in accuracy of the model compared to other models, we can look forward to further development of this model in future related research and applications. For example, in practical applications, the T-GAN model has the potential to serve as an important tool for campus public opinion supervision, accurately identifying negative public opinion, intervening in a timely manner, and ensuring the health of the campus network environment. In addition, its outstanding performance also provides strong reference value for expanding the model to multilingual public opinion analysis in the future.

References

[1] A. P. Pandian, “Performance evaluation and comparison using deep learning techniques in sentiment analysis,” J. Soft Comput. Paradigm, vol. 3, no. 2, pp. 123-134, 2021.

[2] A. Mitra, “Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset),” J. Ubiq. Comput. Commun. Technol., vol. 2, no. 03, pp. 145-152, 2020.

[3] J. Wang, Z. Yang, J. Zhang, Q. Zhang, and W. T. K. Chien, “AdaBalGAN: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition,” IEEE T. Semiconduct. M., vol. 32, no. 3, pp. 310-319, 2019.

[4] S. Kench, and S. J. Cooper, “Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion,” Nat. Mach. Intell., vol. 3, no. 4, pp. 299-305, 2021.

[5] Y. He, L. Zhang, Z. Chen, and C. Y. Li, “A framework of structural damage detection for civil structures using a combined multi-scale convolutional neural network and echo state network,” Eng. Comput., vol. 39, no. 3, pp. 1771-1789, 2023.

[6] M. G. Voskoglou, “A Combined Use of Soft Sets and Grey Numbers in Decision Making,” J. Comput. Cognit. Eng., vol. 2, no. 1, pp. 1-4, 2023.

[7] A. Yadav, and D. K. Vishwakarma, “Sentiment analysis using deep learning architectures: a review,” Artif. Intell. Rev., vol. 53, no. 6, pp. 4335-4385, 2020.

[8] L. Nemes, and A. Kiss, “Social media sentiment analysis based on COVID-19,” J. Inform. Telecommun., vol. 5, no. 1, pp. 1-15, 2021.

[9] P. Mehta, and S. Pandya, “A review on sentiment analysis methodologies, practices and applications,” International Journal of Scientific and Technology Research, vol. 9, no. 2, pp. 601-609, 2020.

[10] S. Kench, and S. J. Cooper, “Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion,” Nat. Mach. Intell., vol. 3, no. 4, pp. 299-305, 2021.

[11] Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, “Generative adversarial networks: A survey toward private and secure applications,” ACM Comput. Surv., vol. 54, no. 6, pp. 1-38, 2021.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, and Y. Bengio, “Generative adversarial networks,” Commun. ACM, vol. 63, no. 11, pp. 139-144, 2020.

[13] R. K. Al-Hamido, “A new neutrosophic algebraic structures,” J. Comput. Cognit. Eng., vol. 2, no. 2, pp. 150-154, 2023.

[14] S. Zhao, L. Yin, J.Zhang, and R. Zhong, “Real-time fabric defect detection based on multi-scale convolutional neural network,” IET Collab. Intell. Manuf., vol. 2, no. 4, pp. 189-196, 2020.

[15] Y.Fang, B. Luo, T. Zhao, D. He, B. Jiang, and Q. Liu, “ST-SIGMA:Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting,” CAAI T. Intell. Techno., vol. 7, no. 4, pp. 744-757, 2022.

[16] L. L. Lemon, and J. Hayes, “Enhancing trustworthiness of qualitative findings: Using Leximancer for qualitative data analysis triangulation,” Qualitat. Rep., vol. 25, no. 3, pp. 604-614, 2020.

[17] A. I. Kadhim, “Survey on supervised machine learning techniques for automatic text classification,” Artif. Intell. Rev., vol. 52, no. 1, pp. 273-292, 2019.

[18] C. Fischer, Z. A. Pardos, R. S. Baker, J. J. Williams, P. Smyth, R. Yu, and M. Warschauer, “Mining big data in education: Affordances and challenges,” Rev. Res. Educ., vol. 44, no. 1, pp. 130-160, 2020.

[19] J. Berger, A. Humphreys, S. Ludwig, W. W. Moe, O. Netzer, and D. A. Schweidel, “Uniting the tribes: Using text for marketing insight,” J. Marketing, vol. 84, no. 1, pp. 1-25, 2020.

[20] M. Wankhade, A. C. S. Rao, and C. Kulkarni, “A survey on sentiment analysis methods, applications, and challenges,” Artif. Intell. Rev., vol. 55, no. 7, pp. 5731-5780, 2022.

[21] Y. He, L. Zhang, Z. Chen, and C. Y. Li, “A framework of structural damage detection for civil structures using a combined multi-scale convolutional neural network and echo state network,” Engineering with Computers, vol. 39, no. 3, pp. 1771-1789, 2023.

[22] S. A. Yazdan, R. Ahmad, N. Iqbal, A. Rizwan, A. N. Khan, and D. H. Kim, “An efficient multi-scale convolutional neural network based multi-class brain MRI classification for SaMD,” Tomography, vol. 8, no. 4, pp. 1905-1927, 2022.

[23] Z. Wen, Y. He, S. Yao, W. Yang, and L. Zhang, “A self-attention multi-scale convolutional neural network method for SAR image despeckling,” International Journal of Remote Sensing, vol. 44, no. 3, pp. 902-923, 2023.

[24] Y. Wang, X. Fan, S. Liu, D. Zhao, and W. Gao, “Multi-scale convolutional neural network-based intra prediction for video coding,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 7, pp. 1803-1815, 2019.

[25] X. Su, J. Xu, Y. Yin, X. Quan, and H. Zhang, “Antimicrobial peptide identification using multi-scale convolutional network,” BMC bioinformatics, vol. 20, no. 1, pp. 1-10, 2019.

[26] K. F. Chu, A. Y. S. Lam, and V. O. K. Li, “Deep multi-scale convolutional LSTM network for travel demand and origin-destination predictions,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 8, pp. 3219-3232, 2019.

[27] P. Zhang, G. Yu, D. Shan, Z. Chen, and X. Wang, “Identifying the Strength Level of Objects’ Tactile Attributes Using a Multi-Scale Convolutional Neural Network,” Sensors, vol. 22, no. 5, pp. 1908-1912, 2022.

[28] C. Tang, X. Liu, X. Zheng, W. Li, L. Wang, and A. Longo, “DeFusionNET: Defocus blur detection via recurrently fusing and refining discriminative multi-scale deep features,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 2, pp. 955-968, 2020.

[29] D. Wu, C. Wang, Y. Wu, Q. C. Wang, and D. S. Huang, “Attention deep model with multi-scale deep supervision for person re-identification,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 5, no. 1, pp. 70-78, 2021.

[30] K T. Chui, B B. Gupta, H R. Chi, V. Arya, W. Alhalabi, M. T. Ruiz, and C. W. Shen, “Transfer learning-based multi-scale denoising convolutional neural network for prostate cancer detection,” Cancers, vol. 14, no. 15, pp. 3687-3691, 2022.

[31] Z. Shen, S P. Deng, and D. S. Huang, “RNA-protein binding sites prediction via multi scale convolutional gated recurrent unit networks,” IEEE/ACM transactions on computational biology and bioinformatics, vol. 17, no. 5, pp. 1741-1750, 2019.

Method for Hyperparameter Tuning of EfficientNetV2-based Image Classification by Deliberately Modifying Optuna Tuned Result

Jin Shimazoe[1], Kohei Arai[2], Mari Oda[3]
Graduate School, Kurume Institute Technology, Kurume City, Japan[1]
Department of Information Science, Saga University, Saga City, Japan[2]
Applied AI Laboratory, Kurume Institute Technology, Kurume City, Japan[2, 3]

***Abstract*—Method for hyperparameter tuning of EfficientNetV2-based image classification by deliberately modifying Optuna tuned result is proposed. An example of the proposed method for textile pattern quality evaluation (good or bad textile pattern fluctuation quality classification) is shown. When using the hyperparameters obtained by Optuna without changing them, the accuracy certainly improved. Furthermore, as a result of learning by changing the hyperparameter with the highest degree of importance, the accuracy changed, so it could be said that the degree of importance was certainly high. However, the accuracy also changes when learning is performed by changing the least important hyperparameter, and sometimes the accuracy is improved compared to when learning is performed using the optimal hyperparameter. From this result, it is found that the optimal hyperparameters obtained with Optuna are not necessarily optimal.**

Keywords—Hyperparameter tuning; EfficientNetV2; Optuna; textile pattern; optimal hyperparameter; learning process; pattern fluctuation

I. Introduction

To optimize these hyperparameters, hyperopt, gpyopt, AutoML, PyCaret, Optuna, etc. have been proposed as black box optimization methods, which automate trial and error regarding hyperparameters and automatically discover optimal solutions. Similarly, as black box optimization methods, white box conversion of DL, binary decision trees, random forests, and mind maps using GNNs has also been proposed [1]. In particular, Optuna uses an algorithm called TPE (Tree-structured Parzen Estimator), which is a new method in Bayesian optimization, and is capable of parallel processing, and can be restarted midway by saving the results to the database.

Depending on the definition of the objective function and the validity of the importance of the parameters, hyperparameters that are not necessarily suitable for comparison with the evaluation criteria may appear. Therefore, in this paper, we introduce such a case and propose a method of intentionally changing the hyperparameters obtained through optimization with Optuna and selecting parameters with greater accuracy through trial and error.

As an application example of this method, we will show an example in which it was applied to the classification of pattern shifts in Kurume Kasuri. This is just one application example, and the proposed method can be widely applied to other classifications.

In Section II, research background and related research works are described followed by the proposed method for hyperparameter tuning by modifying Optuna tuned result in Section III. Then experiment of application of the proposed method given in Section IV followed by remarks in Section V. Conclusion and future research work is given in Section VI and Section VII respectively.

II. Research Background and Related Research Works

A. Research Background

Kurume Kasuri is a traditional cotton fabric handed down in the Chikugo region, and it is completed through over 30 steps, including design, binding, dyeing, and weaving. A major feature of Kurume Kasuri is that the yarn is pre-dyed and the patterned thread is woven while matching the patterns, resulting in subtle deviations and a unique faded pattern.

Regarding the degree of deviation, it is fine if the pattern shift is moderate, but if the deviation is too large, the product will not sell and will have to be sold at a low price. In addition, there is the problem that the evaluation criteria for the degree of deviation differ depending on the manufacturer. Therefore, in this study, we build an image recognition model that classifies whether the pattern shift of Kurume Kasuri is within an acceptable range (good or bad). At that time, Optuna searches for optimal hyperparameters and intentionally changes the most and least important parameters to improve accuracy.

B. Examples of Quality of Kurume Kasuri

Typical patterns of Kurume Kasuri are shown in Fig. 1. This Kurume Kasuri is woven with a rectangular pattern in mind. An example of pattern shift is shown in Fig. 2. Green indicates patterns within the acceptable range, and red indicates patterns outside the acceptable range. The red color has a shape that is almost different from the rectangular pattern, and it can be seen that the pattern is clearly too out of alignment.

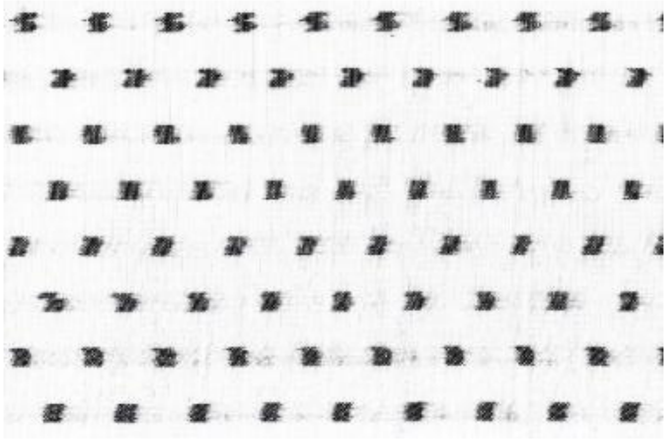


Fig. 1. Typical kurume kasuri image. (sideways)

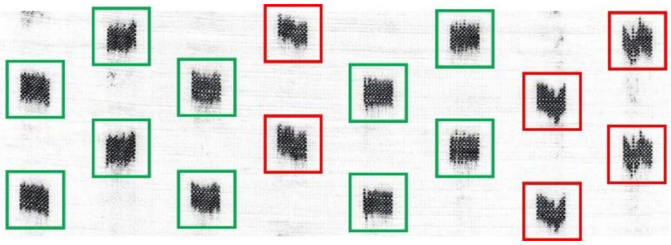


Fig. 2. Example of kurume kasuri pattern shift. (Green: Within the acceptable range, Red: Outside the acceptable range).

C. Related Research Works

It has been proposed a method to evaluate the quality of pattern shifts in Kurume Kasuri by considering them as 1/f fluctuations [2]. Other than this, there are image classification method related research works as follows,

EfficientnetV2: Smaller models and faster training are proposed [3] together with deep neural network configurations of network is network [4].

Classification by re-estimating statistical parameters based on auto-regressive model is proposed [5]. Meanwhile, multi-temporal texture analysis in Landsat Thematic Mapper: TM classification is also proposed [6]. On the other hand, maximum likelihood TM classification taking into account pixel-to-pixel correlation is proposed [7] together with a supervised TM classification with a purification of training samples [8]. Meantime, TM classification using local spectral variability is proposed in [9]. Also, classification method with spatial spectral variability is proposed in [10] together with TM classification using local spectral variability [11].

Application of inversion theory for image analysis and classification methods is proposed [12]. Meanwhile, polarimetric Synthetic Aperture Radar: SAR image classification with maximum curvature of the trajectory in Eigen space domain on the polarization signature is proposed [13].

A hybrid supervised classification method for multi-dimensional images using color and textural features is proposed [14]. On the other hand, polarimetric SAR image classification with high frequency component derived from wavelet Multi Resolution Analysis: MRA is proposed [15].

Comparative study of polarimetric SAR classification methods including proposed method with maximum curvature of trajectory of backscattering cross section in ellipticity and orientation angle space is proposed [16].

Human gait gender classification using 2D discrete wavelet transforms energy is attempted [17] together with human gait gender classification in spatial and temporal reasoning [18]. Meanwhile, comparative study on discrimination methods for identifying dangerous red tide species based on wavelet utilized classification methods is conducted [19].

Multi spectral image classification method with selection of independent spectral features through correlation analysis is proposed [20]. Meanwhile, image retrieval and classification method based on Euclidian distance between normalized features including wavelet descriptor is proposed [21].

Gender classification method based on gait energy motion derived from silhouettes through wavelet analysis of human gait moving pictures is proposed [22] together with human gait skeleton model acquired with single side video camera and its application and implementation for gender classification [23]. Meantime, gender classification method based on gait energy motion derived from silhouette through wavelet analysis of human gait moving pictures is proposed [24] together with human gait gender classification using 3D discrete wavelet transformation feature extraction [25].

Image classification considering probability density function based on Simplified beta distribution is proposed [26]. Maximum likelihood classification based on classified result of boundary mixed pixels for high spatial resolution of satellite images is proposed [27]. On the other hand, context classification based on mixing ratio estimation by means of inversion theory is proposed [28].

Optimum spatial resolution of satellite-based optical sensors for maximizing classification performance is found [29]. Meanwhile, the combined non-parametric and parametric classification method depending on normality of Probability Density Function: PDF of training samples is proposed [30]. In recently, method for hyperparameter tuning of image classification with PyCaret is proposed and well validated [31].

III. Proposed Methods

A. Image Recognition Model

In order to classify whether the pattern shift is within an acceptable range, we used the pre-trained model EfficientNetV2 [3]. EfficientNetV2 is a model that achieves both learning efficiency and high classification accuracy by using NAS (Neural Architecture Search) and model scaling.

Regarding the implementation, using TensorFlow in Python, we added Global Average Pooling [4] and dropout to the final layer of EfficientNetV2, which is a model that has already trained ImageNet, and built a model that changed to binary classification (see Fig. 3). Global Average Pooling is a layer that takes the average value for each feature map obtained in the previous layer. By using this, it is possible to reduce the number of parameters compared to the case of a fully connected layer.

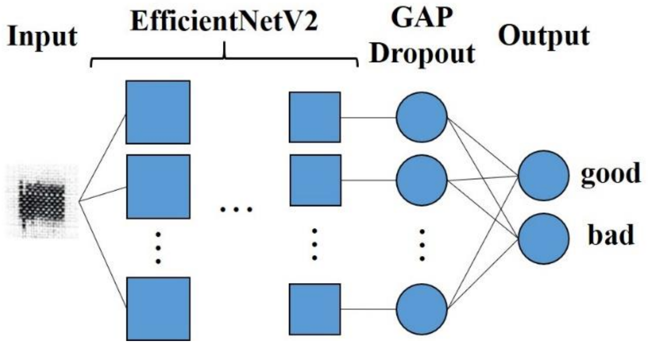


Fig. 3. Model architecture.

With this model, we performed two types of learning: transfer learning only and transfer learning + fine tuning.

B. Hyperparameter Tuning

Hyperparameter tuning was performed using the following three methods, and the accuracy of each method was compared.

1) Manual setting.

2) Optimization using Optuna.

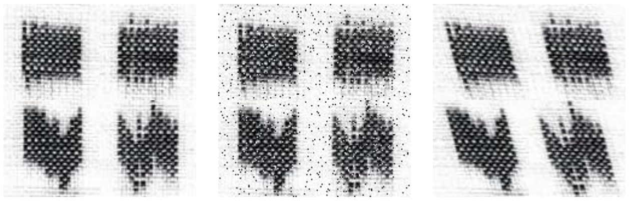
3) Deliberately changing the most important (lowest) hyperparameter among the hyperparameters obtained by Optuna.

In addition, in (2) Optimization using Optuna, specify TPE as the Sampler.

IV. Experiment

A. Data Used

From the scanned Kurume Kasuri images, contour extraction and other operations were performed using OpenCV, and a total of 70 pattern images of 80 × 80 pixels were extracted. Then, based on the results obtained from the pattern evaluation questionnaire to weavers, patterns with pattern shift within the acceptable range were classified as good, and patterns outside the acceptable range were classified as bad [2]. After that, we used a total of 210 image data (training: 180 images, test: 30 images) that was created by applying data augmentation to all of them by adding salt-and-pepper noise and skew (see Fig. 4).



(a) Original (b) Salt and pepper noise (c) Skew

Fig. 4. Sorted Kurume Kasuri pattern image. (upper - good, lower - bad).

B. Transfer Learning

Table I shows the hyperparameters and prediction accuracy for test data when optimizing manually and using Optuna in transfer learning. The two hyperparameters searched were dropout rate and batch size, and the results showed that learning using the optimal hyperparameters obtained by Optuna resulted in better accuracy.

TABLE I. Hyperparameters and Prediction Accuracy for Test Data when Optimized Manually and with Optuna

	Manual	Optuna
Dropout Rate	0.5	0.129 [0 ~ 0.5]
Batch Size	16	32 [16, 32, 64]
Accuracy	76.67%	90%

※The hyperparameter search range is in []

The importance of hyperparameters obtained through optimization using Optuna is shown in Fig. 5. The dropout rate is 0.77 and the batch size is 0.23, indicating that the dropout rate is more important.

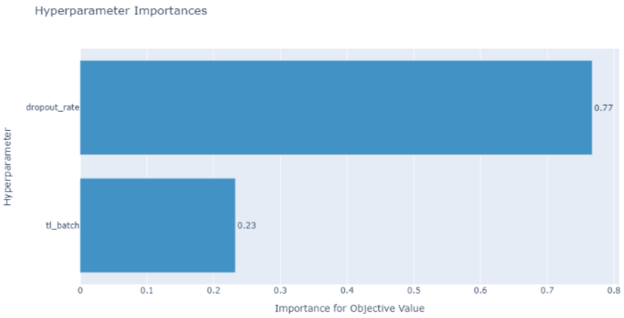
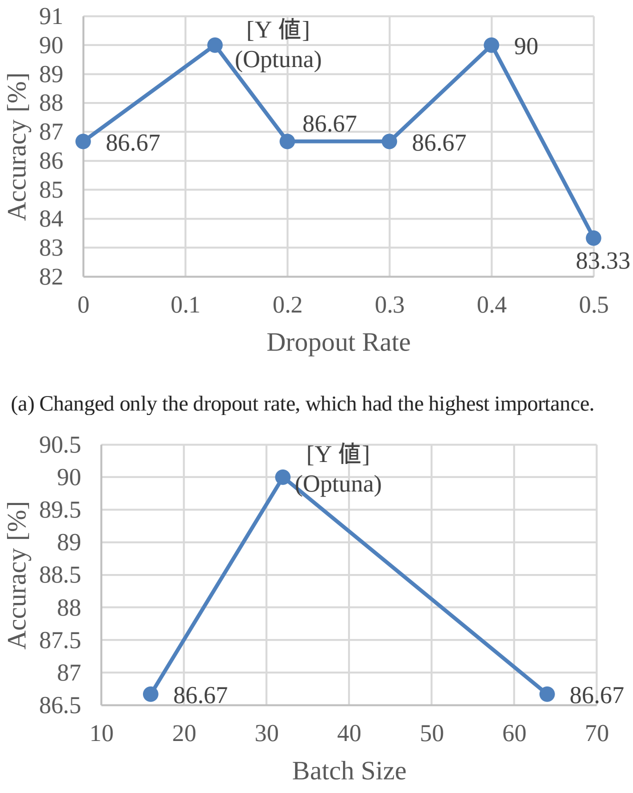


Fig. 5. Importance of hyperparameters when transfer learning was performed.



(a) Changed only the dropout rate, which had the highest importance.

(b) Changed only the batch size, which had the lowest importance.

Fig. 6. Change in accuracy when changing hyperparameters of high (low) importance among the hyperparameters obtained by optimization with Optuna.

Fig. 6(a) shows the change in accuracy when only the dropout rate, which was highly important, was changed. Accuracy changed to some extent, but not regularly, and never exceeded the 90% accuracy in Optuna.

Fig. 6(b) shows the change in accuracy when only changing the batch size, which was of low importance. Accuracy changed little and never exceeded the accuracy of Optuna of 90%, as was the case when only the dropout rate was changed.

C. *Transfer Learning and Fine Tuning*

Table II. shows the hyperparameters and prediction accuracy for test data in transfer learning + fine tuning when optimized manually and with Optuna. The searched hyperparameters were the dropout rate, the learning rate, the number of epochs and batch size for transfer learning, and the batch size for fine tuning.

TABLE II. HYPERPARAMETERS AND PREDICTION ACCURACY FOR TEST DATA WHEN OPTIMIZED MANUALLY AND WITH OPTUNA

	Manual	Optuna
Dropout Rate	0.5	0.124 [0 ~ 0.5]
Learning Rate	0.001	0.001 [0.001, 0.0005, 0.0001]
Epoch (Transfer Learning)	10	15 [10, 15, 20]
Batch Size (Transfer Learning)	16	32 [16, 32]
Batch Size (Fine Tuning)	16	32 [16, 32]
Accuracy	50%	80%

※The hyperparameter search range is in []

As with the case where only transfer learning was performed, the optimal hyperparameters obtained by Optuna were used. The result was that the accuracy was better when learning was performed. The importance of hyperparameters obtained through optimization using Optuna is shown in Fig. 7. The dropout rate was the most important at 0.49, and the batch size (transfer learning) was the lowest at 0.05.

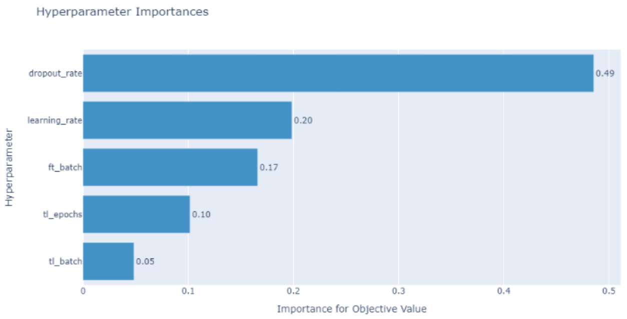
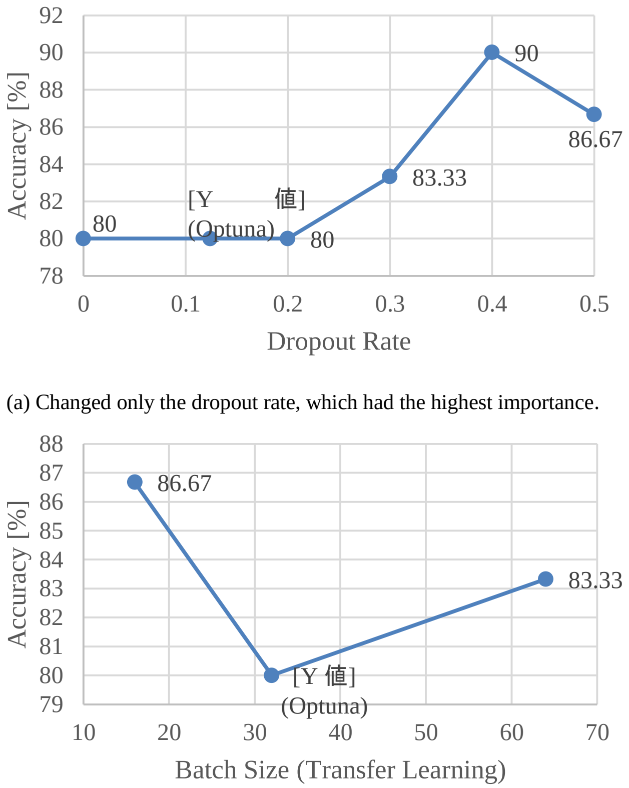


Fig. 7. Importance of hyperparameters when transfer learning and fine tuning were performed.

Fig. 8(a) shows the change in accuracy when only the dropout rate, which was highly important, was changed. Unlike the case of only transfer learning, the accuracy changed significantly, and when the dropout rate was 0.4, the accuracy was 90%, which exceeded the accuracy of Optuna of 80%.

Fig. 8(b) shows the change in accuracy when only changing the batch size (transfer learning), which was of low importance. Unlike the case of only transfer learning, the accuracy changed to some extent, and exceeded the accuracy of Optuna of 80%, which is the same as when changing only the dropout rate.



(a) Changed only the dropout rate, which had the highest importance.

(b) Changed only the batch size in the transfer learning, which had the lowest importance.

Fig. 8. Change in accuracy when changing hyperparameters of high (low) importance among the hyperparameters obtained by optimization with optuna.

V. REMARKS

In both the case of transfer learning only and the case of performing fine tuning after transfer learning, the dropout rate was the most important. This is thought to be because only the dropout rate had a continuous search range and the widest search range, while the other hyperparameters had a categorical search range.

The reason why the accuracy change was not regular when only the dropout rate was changed is because the nodes deleted by Dropout are random, so if a node with features that have a large impact on classification is deleted. This is thought to be due to the fact that there were cases where this was not done. In order to measure accuracy more accurately, it is necessary to perform verification using K-fold cross validation.

In transfer learning + fine tuning, by changing the hyperparameters with the highest (lowest) importance, the accuracy was improved compared to the results with Optuna. This is because Optuna's search algorithm used this time, TPE, is based on Bayesian optimization and does not exhaustively search all hyperparameters like grid search, so a locally optimal solution was reached. This is thought to be the cause.

VI. CONCLUSION

We proposed a method of intentionally changing the hyperparameters obtained through optimization using Optuna and selecting parameters with greater accuracy through trial and error. As an application of the proposed method, we classified pattern shifts in Kurume Kasuri.

In both cases, whether it's transfer learning alone or fine-tuning after transfer learning, learning with hyperparameters obtained through optimization with Optuna clearly improved accuracy compared to setting them manually. In transfer learning + fine tuning, by changing the hyperparameters with the highest (lowest) importance, the accuracy was improved compared to the results with Optuna.

From this result, we found that the optimal hyperparameters obtained with Optuna are not necessarily optimal.

VII. FUTURE RESEARCH WORKS

Further study is required for validation of the proposed method for hyperparameter tuning with a variety of examples of image classifications. Also, the other method for optimization of automate hyperparameter search has to be investigated in the near future.

ACKNOWLEDGMENT

The authors would like to thank the team members of Applied AI laboratory of Kurume Institute of Technology for their useful discussions we have had through this study.

REFERENCES

[1] Kohei Arai, Method for Training and White Boxing DL, BDT, Random Forest and Mind Maps Based on GNN, Appl. Sci. 2023, 13, 4743. https://doi.org /10.3390/app13084743/, 2023.

[2] Jin Shimazoe, Kohei Arai, Mariko Oda, Jewon Oh, Method for 1/f Fluctuation Component Extraction from Images and Its Application to Improve Kurume Kasuri Quality Estimation, International Journal of Advanced Computer Science and Applications, 13, 11, 465-471, 2022.

[3] Mingxing Tan, Quoc Le, EfficientnetV2: Smaller models and faster training, International Conference on Machine Learning, PMLR 139:10096-10106, 2021.

[4] Min Lin, Qiang Chen, Shuicheng Yan, Network in network, arXiv preprint arXiv:1312.4400, 2013.

[5] Kohei Arai, Classification by Re-Estimating Statistical Parameters Based on Auto-Regressive Model, Canadian Journal of Remote Sensing, Vol.16, No.3, pp.42-47, Jul.1990.

[6] Kohei Arai, Multi-Temporal Texture Analysis in TM Classification, Canadian Journal of Remote Sensing, Vol.17, No.3, pp.263-270, Jul.1991.

[7] Kohei Arai, Maximum Likelihood TM Classification Taking into account Pixel-to-Pixel Correlation, Journal of International GEOCARTO, Vol.7, pp.33-39, Jun.1992.

[8] Kohei Arai, A Supervised TM Classification with a Purification of Training Samples, International Journal of Remote Sensing, Vol.13, No.11, pp.2039-2049, Aug.1992.

[9] Kohei Arai, TM Classification Using Local Spectral Variability, Journal of International GEOCARTO, Vol.7, No.4, pp.1-9, Oct.1992.

[10] Kohei Arai, A Classification Method with Spatial Spectral Variability, International Journal of Remote Sensing, Vol.13, No.12, pp.699-709, Oct.1992.

[11] Kohei Arai, TM Classification Using Local Spectral Variability, International Journal of Remote Sensing, Vol.14, No.4, pp.699-709, 1993.

[12] Kohei Arai, Application of Inversion Theory for Image Analysis and Classification, Advances in Space Research, Vol.21, 3, 429-432, 1998.

[13] Kohei Arai and J.Wang, Polarimetric SAR image classification with maximum curvature of the trajectory in eigen space domain on the polarization signature, Advances in Space Research, 39, 1, 149-154, 2007.

[14] Hiroshi Okumura, Makoto Yamaura and Kohei Arai, A hybrid supervised classification method for multi-dimensional images using color and textural features, Journal of the Institute of Image Electronics Engineers of Japan、38, 6, 872-882, 2009.

[15] Kohei Arai, Polarimetric SAR image classification with high frequency component derived from wavelet multi resolution analysis: MRA, International Journal of Advanced Computer Science and Applications, 2, 9, 37-42, 2011.

[16] Kohei Arai Comparative study of polarimetric SAR classification methods including proposed method with maximum curvature of trajectory of backscattering cross section in ellipticity and orientation angle space, International Journal of Research and Reviews on Computer Science, 2, 4, 1005-1009, 2011.

[17] Kohei Arai, Rosa Andrie, Human gait gender classification using 2D discrete wavelet transforms energy, International Journal of Computer Science and Network Security, 11, 12, 62-68, 2011.

[18] Kohei Arai, R.A.Asunara, Human gait gender classification in spatial and temporal reasoning, International Journal of Advanced Research in Artificial Intelligence, 1, 6, 1-6, 2012.

[19] Kohei Arai, Comparative study on discrimination methods for identifying dangerous red tide species based on wavelet utilized classification methods, International Journal of Advanced Computer Science and Applications, 4, 1, 95-102, 2013.

[20] Kohei Arai, Multi spectral image classification method with selection of independent spectral features through correlation analysis, International Journal of Advanced Research in Artificial Intelligence, 2, 8, 21-27, 2013.

[21] Kohei Arai, Image retrieval and classification method based on Euclidian distance between normalized features including wavelet descriptor, International Journal of Advanced Research in Artificial Intelligence, 2, 10, 19-25, 2013.

[22] Kohei Arai, Rosa Andrie Asmara, Gender classification method based on gait energy motion derived from silhouettes through wavelet analysis of human gait moving pictures, International Journal of Information Technology and Computer Science, 6, 3, 1-11, 2014.

[23] Kohei Arai, Rosa Andrie Asmara, Human gait skeleton model acquired with single side video camera and its application and implementation for gender classification, Journal of the Image Electronics and Engineering Society of Japan, Transaction of Image Electronics and Visual Computing, 1, 1, 78-87, 2013.

[24] Kohei Arai, Rosa Andrie Asmara, Gender classification method based on gait energy motion derived from silhouette through wavelet analysis of human gait moving pictures, International Journal of Information technology and Computer Science, 5, 5, 12-17, 2013.

[25] Kohei Arai, Rosa Andrie Asmara, Human gait gender classification using 3D discrete wavelet transformation feature extraction, International Journal of Advanced Research in Artificial Intelligence, 3, 2, 12-17, 2014.

[26] Kohei Arai, Image classification considering probability density function based on Simplified beta distribution, International Journal of Advanced Computer Science and Applications IJACSA, 11, 4, 481-486, 2020.

[27] Kohei Arai, Maximum Likelihood Classification based on Classified Result of Boundary Mixed Pixels for High Spatial Resolution of Satellite Images, International Journal of Advanced Computer Science and Applications, Vol. 11, No. 9, 24-30, 2020.

[28] Kohei Arai, Context Classification based on Mixing Ratio Estimation by Means of Inversion Theory, International Journal of Advanced Computer Science and Applications, Vol. 11, No. 12, 44-50, 2020.

[29] Kohei Arai, Optimum Spatial Resolution of Satellite-based Optical Sensors for Maximizing Classification Performance, International Journal of Advanced Computer Science and Applications, Vol. 12, No. 2, 363-369, 2021.

[30] Kohei Arai, Combined Non-parametric and Parametric Classification Method Depending on Normality of PDF of Training Samples, International Journal of Advanced Computer Science and Applications, Vol. 12, No. 5, 310-316, 2021.

[31] Kohei Arai, Jin Shimazoe, Mariko Oda, Method for Hyperparameter Tuning of Image Classification with PyCaret, International Journal of Advanced Computer Science and Applications, Vol. 14, No. 9, 276-282, 2023.

AUTHORS' PROFILE

Jin Shimazoe, He received BE degree in 2022. He also received the IEICE Kyushu Section Excellence Award. He is currently working on research that uses image processing and image recognition in Master's Program at Kurume Institute of Technology.

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR during 2008 and 2020 then he is now award committee member of ICSU/COSPAR. He is now Visiting Professor of Nishi-Kyushu University since 2021, and is Visiting Professor of Kurume Institute of Technology (Applied AI Laboratory) since 2021. He wrote 87 books and published 700 journal papers as well as 570 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html.

Mariko Oda, She graduated from the Faculty of Engineering, Saga University in 1992, and completed her master's and doctoral studies at the Graduate School of Engineering, Saga University in 1994 and 2012, respectively. She received Ph.D (Engineering) from Saga University in 2012. She also received the IPSJ Kyushu Section Newcomer Incentive Award. In 1994, she became an assistant professor at the department of engineering in Kurume Institute of Technology; in 2001, a lecturer; from 2012 to 2014, an associate professor at the same institute; from 2014, an associate professor at Hagoromo university of International studies; from 2017 to 2020, a professor at the Department of Media studies, Hagoromo university of International studies. In 2020, she was appointed Deputy Director and Professor of the Applied of AI Research Institute at Kurume Institute of Technology. She has been in this position up to the present. She is currently working on applied AI research in the fields of education.

A Novel Framework for Risk Prediction in the Health Insurance Sector using GIS and Machine Learning

Prasanta Baruah, Pankaj Pratap Singh, Sanjiv kumar Ojah
Department of Computer Science & Engineering, Central Institute of Technology Kokrajhar, Kokrajhar, India

***Abstract*—Evaluation of risk is a key component to categorize the customers of the life insurance businesses. The underwriting technique is carried out by the industries to charge the policies appropriately. Due to the availability of data hugely, the automation of underwriting process can be done using data analytics technology. Due to this, the underwriting process becomes faster and therefore quickly processes a large number of applications. This study is carried to enhance risk assessment of the applicants of life insurance industries using predictive analytics. In this research, the Geographical Information Systems (GIS) system is used to collect the data such as Air pollution, Industrial area, Covid-19 and Malaria of various geographic areas of our country, since these factors attribute to the risk of an applicant of life insurance business. Thereafter, the research is carried out using this dataset along with another dataset containing more than 50,000 entries of normal attributes of applicants of a life insurance company. Artificial Neural Network (ANN), Decision Tree (DT), and Random forest (RF) algorithms are applied on both the datasets to predict the risks of the applicants. The results showed that random forest outperformed among all the algorithms, providing the more accurate result.**

***Keywords*—Risk prediction; data analytics; predictive analytics; underwriting; geographical information systems; random forest; artificial neural network; decision tree**

I. Introduction

Assessment of risk plays a significant role in the classification of applicants in any life insurance industry. The life insurance businesses use underwriting procedure to choose applications and set insurance product prices accordingly. For quicker application processing, the underwriting procedure might be automated thanks to the growth of data and developments in data analytics [13]. In order to develop solutions that cater to the demands of various client and market groups, insurance companies are giving emphasis of using machine learning (ML) techniques on big data. The evaluation comprises estimating the company's risk factor and offering potential employees health insurance based on their medical histories.

Developing a scalable risk assessment model for the customer segment of the insurance domain using geospatial technologies alike Global Positioning System (GPS) is very much suitable, particularly for surveillance of areas appearing infectious disease or environmental health hazards such as air pollution, water pollution, etc.,. This research discusses how the insurance companies can use a GIS data model to analyze these type parameters across the globe while addressing their biggest bottleneck. Numerous tools and systems will be developed that enable the visualization of disease/health hazard data in location and time as a result of this ongoing public health burden and technological advancements with spatial data [5]. This geospatial technology will therefore offer insurance businesses and decision maker’s visualization and analytical tools to conduct life insurance programs for clients in afflicted and/or suspected locations as well as analysis and forecasts that were previously technologically impractical.

The research problem addressed in this study is the development of an innovative framework leveraging Geographic Information Systems (GIS) and Machine Learning techniques for enhanced risk prediction in the Health Insurance Sector. The research question comes like as “What specific methodologies and algorithms within GIS and Machine Learning can enhance health insurance risk prediction?” The primary objectives of this study are to design and implement a novel GIS and Machine Learning framework for accurate health insurance risk prediction, assess its performance, and provide recommendations for practical implementation in the health insurance sector. The research holds significance by offering an innovative approach to health insurance risk prediction, potentially improving accuracy and decision-making in the sector through the integration of GIS and Machine Learning technologies. The main research contribution is the development of an advanced and novel framework that integrates GIS and Machine Learning to enhance the accuracy and effectiveness of risk prediction in the Health Insurance Sector.

GIS, GPS, and satellite-based technologies such as Remote Sensing (RS) are all examples of geospatial technology. GIS refers to the collection, input, updating, modification, transformation, analysis, query, modeling, and visualization of geographic data utilizing a collection of computer programs. The next Section II and Section III will discuss about the related research work and methodology respectively. The detailed results related discussion is mentioned in Section IV and finally conclusion given in Section V followed by the references section of this paper.

II. Related Work

A map-based dashboard was proposed for visualizing the COVID19 pandemic in order to deliver information to individuals all around the world who want to safeguard themselves and their communities and how a complete GIS platform may aid in the surveillance, preparedness, and response of infectious diseases [3]. By collecting data from satellites can be made available to users. Satellites are terrain surveillance equipment that provides regional information on climatic parameters and terrain features. Additionally, GPS uses satellite data to provide locating, direction-finding, and timing services. As a result, while GPS and RS provide local and spatial information, GIS offers accurate geospatial analysis and real-time geospatial data integration. This study is conducted and found that machine learning algorithms like DT, RF, and ANN are effective in estimating the risk level of applicants in an insurance industry [4, 7].

In this paper, a potential risk variables was investigated those contribute to the cases of COVID-19 at the various districts of Bangladesh. In this work, three global models and one local model were built based on demographic, economic, built environment along with the factors like health and facilities which affect rates of COVID-19 occurrence cases [10]. It was found that the percentage of urban population of the districts was responsible for the COVID-19 occurrence rates. This is due to the fact that in high-density urban regions, movements of people as well as activities are higher than the non-urban regions. The researchers discovered that the higher the inhabitant's density, there is more possibility that a person will come into contact with an infector. In this paper, a framework is presented which shows the different flood hazards in spatial hotspots areas and also assessed the vulnerability in the districts using MODIS data [16]. A ML approach was used to classify and find risk based on diabetics disease [17].

In the univariate analysis, population density was also revealed to be a significant variable in this study [2]. In this paper, authors found that in the OCHIN (Oregon Community Health Information Network) PBRN (practice-based research network) consist of community health centres. These networks were used to display of EHR (electronic health record) data using GIS web based mapping and thereby serving in detecting societies having higher number of patients without health insurance. The author suggested that this strategy might be adapted for use by PBRNs, primary care providers, public health officials, and others to recognize a wide range of practice and community needs and to correctly implement targeted interventions using additional EHR data pieces [5]. This paper work added current COVID-19 assessments by providing a geographical viewpoint. It's a collection that recognizes the themes and analyses that GIS and spatial-statistical tools are being used for [1]. The detailed comparative studies are mentioned below in the Table I which is done by the several researchers.

The main objective of this research work is to develop a software model that makes use of web GIS technology to calculate risk in a specific location using information for the health insurance industry. As a result, the insurance company might carry out a more thorough examination and come to better judgments that will benefit both their clients and the business. The insurance company may compare data from the previous year to help it make better decisions. Additionally, with the use of various hazard data, such as information on air pollution, malaria, COVID-19, industrial regions, etc., the companies may analyze the geographic area and make better decisions in order to provide consumers a variety of life insurance plans.

TABLE I. COMPARATIVE STUDY OF THE ALGORITHMS USED IN THE RELATED WORKS

Authors	Year	Objective	Methods used	Method giving the best performance
Rahman et al. [2]	2021	To identify the risk factors of Covid-19	GIS based spatial model	GIS model
Saran et al. [6]	2020	Reviewing of Geospatial Technology	GIS approach and dynamic modelling algorithms	GIS approach
Boulos et al. [3]	2020	Tracing and mapping of Covid-19 patients	GIS system	GIS system
Pardo et al. [1]	2020	Covid-19 analysis	GIS and spatial analysis	GIS method
Angier et al. [5]	2014	Identification of communities in need of health insurance	GIS web based	GIS web based
Boodhun et al. [4]	2018	Implementation of ML algorithms to classify the applicants risk levels	Multiple Linear Regression (MLR), ANN, RF, REPTree algorithms	REPTree performed better with Correlation Based Feature Selection (CFS) whereas MLR showed the best performance using Principal Components Analysis (PCA) method
Hutagaol et al. (2020) [11]		Examined risk level of customers using ML in life insurance companies	RF, Support Vector Machine (SVM) and Naive Bayesian algorithms	RF have highest precision in comparison to the SVM and Naive Bayesian (N.B.) algorithm
Mustika et al. [14]	2019	Applied a ML models to predict the risk level of applicants in life insurance	Extreme Gradient tree boosting (XGBoost), DT, RF and Bayesian ridge models	XGBoost model is provided more accurate result
Jain et al. [12]	2019	An ensemble learning method for assessing risk associated with a policy applicant	ANN and gradient boosting algorithm XGBoost	XGBoost provided the best result.
Biddle et al. [13]	2018	To automate the underwriting process	Logistic Regression, XGBoost and Recursive Feature Elimination	XGBoost is the most ideal one giving better accurac.
Dwivedi et. al. [15]	2020	ML algorithms are used to predict the risk levels of applicants	ANN, MLR, RT and RF	RF came out to be most efficient one.

III. Methodology

As part of the research methodology, data is acquired from online databases. The research paradigm adopts a positivist stance because the study is largely predictive and uses machine learning and some geospatial approaches to assist this research work goals. The major task of the gathered data is to use Quantum-GIS (QGIS) to turn all of the non-spatial data into spatial data. The process flow of the model which utilizes different techniques is shown in the Fig. 1. The details of the data, format, source and description are mentioned in the given below Table II.

A. Data Collection and Preparation

The 59,381 applications that make up the life insurance data collection each have 128 attributes that describe the traits of applicants for life insurance. The data set contains anonymised nominal, continuous, and discrete variables. Customer related sample dataset of the applicants is shown in the Table III. The data pre-processing, commonly referred to as data cleaning, which comprises removing erratic data or outliers from the target dataset. This step also includes developing any methods necessary to deal with the target data's inconsistencies. To facilitate analysis and interpretation in the event of disputes, certain variables will be changed. In this phase, the data will be cleaned to get rid of any missing values and make sure it can be used for the analysis. To determine the optimal imputation procedure for the dataset, it will be necessary to look into the structure and methodology of missing data.

B. Geospatial Technology

Geospatial technologies and services can help with the representation of spatio-temporal data and the analysis of dynamic spread when illnesses or hazards are present [8]. A "spatial health information infrastructure" must be built using geospatial technology and real time services. In order to better comprehend analysis the spread and outbreak of the phenomenon, a review of several geospatial technologies with enabled IT services will be conducted in this section using a case study on the COVID-19 pandemic, malaria illness, air pollution, and industrial region.

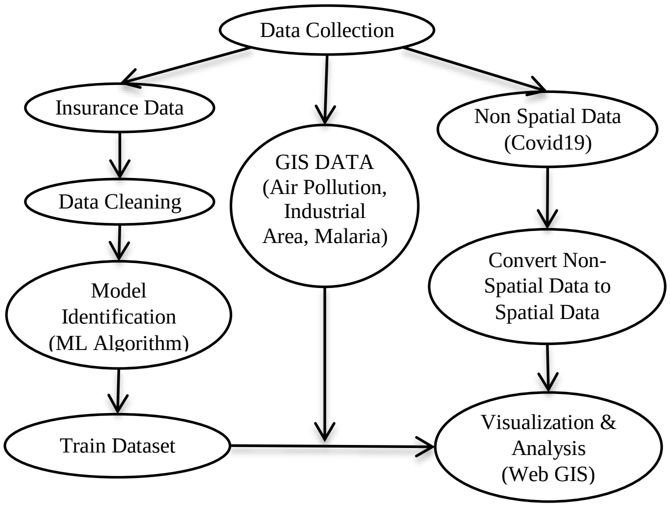

Data Collection
Insurance Data
GIS DATA (Air Pollution, Industrial Area, Malaria)
Non Spatial Data (Covid19)
Data Cleaning
Convert Non-Spatial Data to Spatial Data
Model Identification (ML Algorithm)
Train Dataset
Visualization & Analysis (Web GIS)

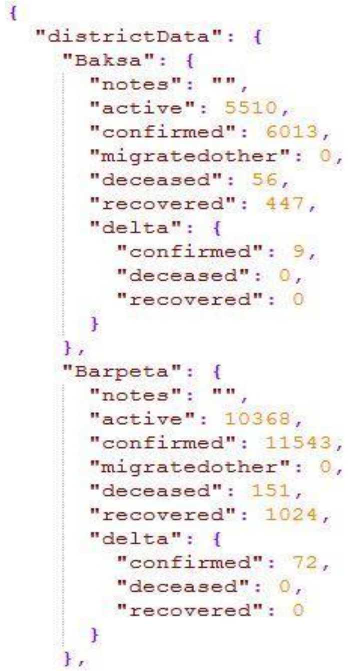
Fig. 1. Flowchart of the system.

TABLE II. Comparative Data Source and Description

Data	Source and Information	Format
Air Quality	MODIS	Raster Data
Industrial Area	NESDR	Vector Data
Covid19	covid19india.org	JSON
Malaria	NESDR	GEOJSON

TABLE III. Sample Dataset of Customer

Id	Product_Info_1	Ins_Age	Ht	Wt	BMI	Employment_Info_1
2	1	0.641791	0.581818	0.148536	0.323008	0.028
5	1	0.059701	0.6	0.131799	0.272288	0
6	1	0.029851	0.745455	0.288703	0.42878	0.03
7	1	0.164179	0.672727	0.205021	0.352438	0.042
8	1	0.41791	0.654545	0.23431	0.424046	0.027
10	1	0.507463	0.836364	0.299163	0.364887	0.325
11	1	0.373134	0.581818	0.17364	0.376587	0.11
14	1	0.61194	0.781818	0.403766	0.571612	0.12
15	1	0.522388	0.618182	0.1841	0.362643	0.165
16	1	0.552239	0.6	0.284519	0.587796	0.025
17	1	0.537313	0.690909	0.309623	0.521668	0.05
18	1	0.298507	0.690909	0.271967	0.45505	0.09
19	1	0.567164	0.618182	0.16318	0.320784	0.075
20	2	0.223881	0.781818	0.361925	0.507515	0.1
22	1	0.328358	0.636364	0.142259	0.264648	0.16
23	1	0.626866	0.672727	0.330544	0.581279	0.075



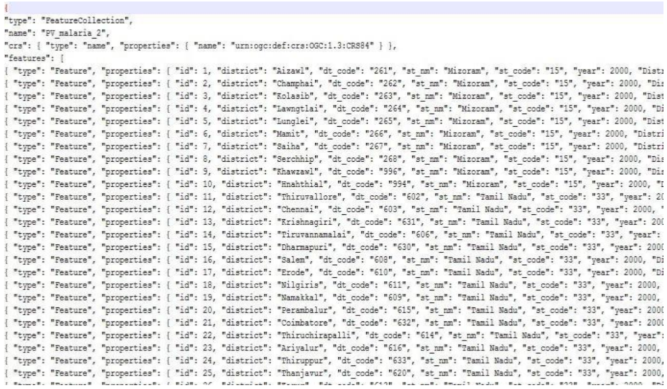
```
{
  "districtData": {
    "Baksa": {
      "notes": "",
      "active": 5510,
      "confirmed": 6013,
      "migratedother": 0,
      "deceased": 56,
      "recovered": 447,
      "delta": {
        "confirmed": 9,
        "deceased": 0,
        "recovered": 0
      }
    },
    "Barpeta": {
      "notes": "",
      "active": 10368,
      "confirmed": 11543,
      "migratedother": 0,
      "deceased": 151,
      "recovered": 1024,
      "delta": {
        "confirmed": 72,
        "deceased": 0,
        "recovered": 0
      }
    },
```

Fig. 2. COVID-19 sample data in JSON format.

Fig. 2 shows the sample data of covid-19 for the districts of Assam in JSON format. Fig. 3 shows the sample data of malaria-PV (Plasmodium Vivax) virus and Fig. 4 shows the sample data of malaria-PF (Plasmodium Falciparum) virus in GEOJSON format of all the districts of India [9]. The red marks are shown in the Fig. 5 which denotes the industrial areas of NER for the 3rd Land Use Land Cover (LULC) cycle (2015-2016). Fig. 6 shows the MODIS data of NER air pollution for 15 January 2021.

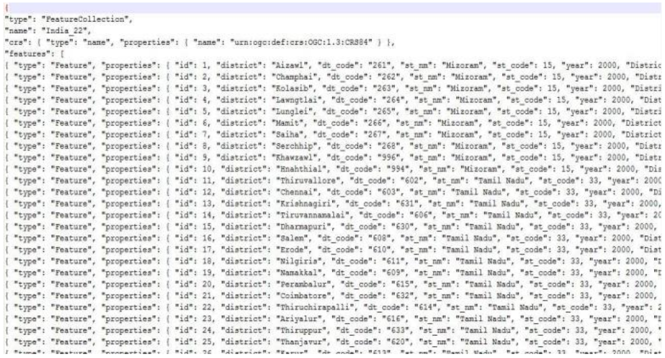
C. Model Development

The optimal model is chosen using a combination of decision tree and ensemble techniques because risk level is a multi-class variable. Combining data from a number of different models to increase accuracy and stability is referred to as an "ensemble." The many algorithms that were used to create the predictive models using the data set were covered in this section. Several machine learning methods, including DT, RF, and ANN have been applied on the dataset after pre-processing the data.



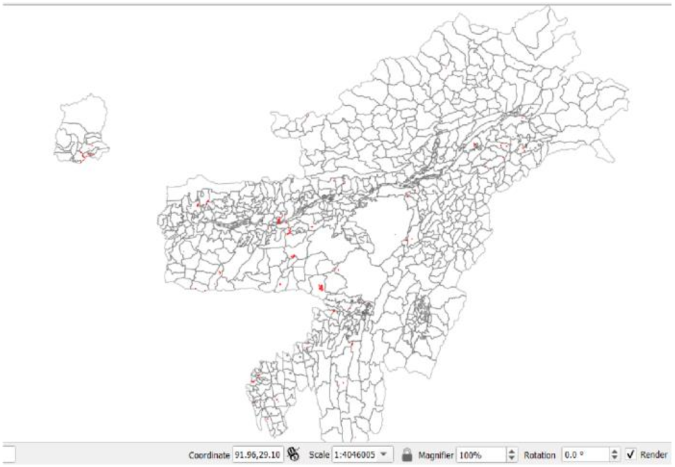
```
{
"type": "FeatureCollection",
"name": "PV_malaria_2",
"crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:OGC:1.3:CRS84" } },
"features": [
{ "type": "Feature", "properties": { "id": 1, "district": "Aizawl", "dt_code": "261", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Distr
{ "type": "Feature", "properties": { "id": 2, "district": "Champhai", "dt_code": "262", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Dis
{ "type": "Feature", "properties": { "id": 3, "district": "Kolasib", "dt_code": "263", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Dist
{ "type": "Feature", "properties": { "id": 4, "district": "Lawngtlai", "dt_code": "264", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Di
{ "type": "Feature", "properties": { "id": 5, "district": "Lunglei", "dt_code": "265", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Dist
{ "type": "Feature", "properties": { "id": 6, "district": "Mamit", "dt_code": "266", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Distri
{ "type": "Feature", "properties": { "id": 7, "district": "Saiha", "dt_code": "267", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Distri
{ "type": "Feature", "properties": { "id": 8, "district": "Serchhip", "dt_code": "268", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Dis
{ "type": "Feature", "properties": { "id": 9, "district": "Khawzawl", "dt_code": "996", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "Dis
{ "type": "Feature", "properties": { "id": 10, "district": "Hnahthial", "dt_code": "994", "st_nm": "Mizoram", "st_code": "15", "year": 2000, "I
{ "type": "Feature", "properties": { "id": 11, "district": "Thiruvallore", "dt_code": "602", "st_nm": "Tamil Nadu", "st_code": "33", "year": 20
{ "type": "Feature", "properties": { "id": 12, "district": "Chennai", "dt_code": "603", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000, "
{ "type": "Feature", "properties": { "id": 13, "district": "Krishnagiri", "dt_code": "631", "st_nm": "Tamil Nadu", "st_code": "33", "year": 200
{ "type": "Feature", "properties": { "id": 14, "district": "Tiruvannamalai", "dt_code": "606", "st_nm": "Tamil Nadu", "st_code": "33", "year":
{ "type": "Feature", "properties": { "id": 15, "district": "Dharmapuri", "dt_code": "630", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000
{ "type": "Feature", "properties": { "id": 16, "district": "Salem", "dt_code": "608", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000, "Di
{ "type": "Feature", "properties": { "id": 17, "district": "Erode", "dt_code": "610", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000, "Di
{ "type": "Feature", "properties": { "id": 18, "district": "Nilgiris", "dt_code": "611", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000,
{ "type": "Feature", "properties": { "id": 19, "district": "Namakkal", "dt_code": "609", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000,
{ "type": "Feature", "properties": { "id": 20, "district": "Perambalur", "dt_code": "615", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000
{ "type": "Feature", "properties": { "id": 21, "district": "Coimbatore", "dt_code": "632", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000
{ "type": "Feature", "properties": { "id": 22, "district": "Thiruchirapalli", "dt_code": "614", "st_nm": "Tamil Nadu", "st_code": "33", "year":
{ "type": "Feature", "properties": { "id": 23, "district": "Ariyalur", "dt_code": "616", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000,
{ "type": "Feature", "properties": { "id": 24, "district": "Thiruppur", "dt_code": "633", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000,
{ "type": "Feature", "properties": { "id": 25, "district": "Thanjavur", "dt_code": "620", "st_nm": "Tamil Nadu", "st_code": "33", "year": 2000,
```

Fig. 3. Malaria- PV (Plasmodium Vivax) sample data in GEOJSON format.



```
{
"type": "FeatureCollection",
"name": "India_22",
"crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:OGC:1.3:CRS84" } },
"features": [
{ "type": "Feature", "properties": { "id": 1, "district": "Aizawl", "dt_code": "261", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "Distric
{ "type": "Feature", "properties": { "id": 2, "district": "Champhai", "dt_code": "262", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "Distr
{ "type": "Feature", "properties": { "id": 3, "district": "Kolasib", "dt_code": "263", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "Distri
{ "type": "Feature", "properties": { "id": 4, "district": "Lawngtlai", "dt_code": "264", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "Dist
{ "type": "Feature", "properties": { "id": 5, "district": "Lunglei", "dt_code": "265", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "Distri
{ "type": "Feature", "properties": { "id": 6, "district": "Mamit", "dt_code": "266", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "District
{ "type": "Feature", "properties": { "id": 7, "district": "Saiha", "dt_code": "267", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "District
{ "type": "Feature", "properties": { "id": 8, "district": "Serchhip", "dt_code": "268", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "Distr
{ "type": "Feature", "properties": { "id": 9, "district": "Khawzawl", "dt_code": "996", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "Distr
{ "type": "Feature", "properties": { "id": 10, "district": "Hnahthial", "dt_code": "994", "st_nm": "Mizoram", "st_code": 15, "year": 2000, "Dis
{ "type": "Feature", "properties": { "id": 11, "district": "Thiruvallore", "dt_code": "602", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000
{ "type": "Feature", "properties": { "id": 12, "district": "Chennai", "dt_code": "603", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000, "Di
{ "type": "Feature", "properties": { "id": 13, "district": "Krishnagiri", "dt_code": "631", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000,
{ "type": "Feature", "properties": { "id": 14, "district": "Tiruvannamalai", "dt_code": "606", "st_nm": "Tamil Nadu", "st_code": 33, "year": 20
{ "type": "Feature", "properties": { "id": 15, "district": "Dharmapuri", "dt_code": "630", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000,
{ "type": "Feature", "properties": { "id": 16, "district": "Salem", "dt_code": "608", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000, "Dist
{ "type": "Feature", "properties": { "id": 17, "district": "Erode", "dt_code": "610", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000, "Dist
{ "type": "Feature", "properties": { "id": 18, "district": "Nilgiris", "dt_code": "611", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000, "I
{ "type": "Feature", "properties": { "id": 19, "district": "Namakkal", "dt_code": "609", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000, "I
{ "type": "Feature", "properties": { "id": 20, "district": "Perambalur", "dt_code": "615", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000,
{ "type": "Feature", "properties": { "id": 21, "district": "Coimbatore", "dt_code": "632", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000,
{ "type": "Feature", "properties": { "id": 22, "district": "Thiruchirapalli", "dt_code": "614", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2
{ "type": "Feature", "properties": { "id": 23, "district": "Ariyalur", "dt_code": "616", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000, "I
{ "type": "Feature", "properties": { "id": 24, "district": "Thiruppur", "dt_code": "633", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000, "
{ "type": "Feature", "properties": { "id": 25, "district": "Thanjavur", "dt_code": "620", "st_nm": "Tamil Nadu", "st_code": 33, "year": 2000, "
```

Fig. 4. Malaria- PF (Plasmodium Falciparum) sample data in GEOJSON format.



Coordinate 91.96,29.10 Scale 1:4046005 Magnifier 100% Rotation 0.0 ° ✓ Render

Fig. 5. Industrial area depiction in NER region using QGI software environment.

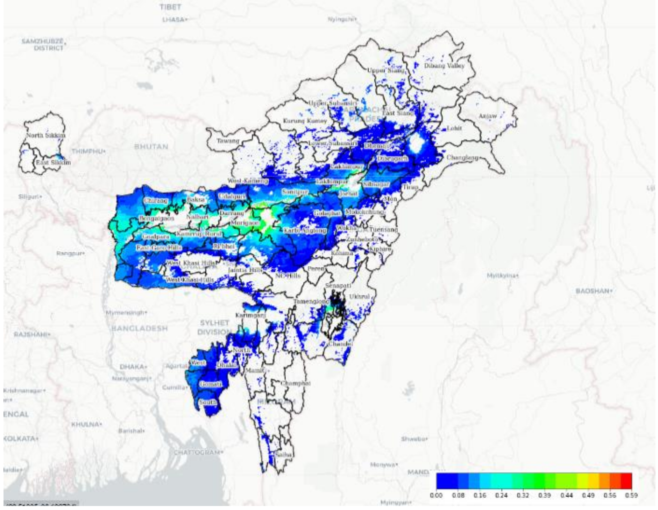


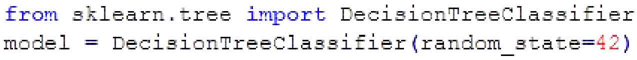
Fig. 6. MODIS data of NER air pollution.

1) Decision Tree: The SKlearn module in Python is used to build the DT as shown in Fig. 7 and also the information gain is calculated with the help of sample estimation which are shown below in Eq. (1), (2) and (3). The accuracy of this algorithm is coming 80%.

$$\mathrm{E(s)} = \sum_{i=1}^{c}(-p_i \log_2 p_i) \qquad (1)$$

$$\mathrm{E(T, X)} = \sum_{c\in X} P(c)E(c) \qquad (2)$$

$$\mathrm{InforGain\ (T, X) = E(T) - E(T, X)} \qquad (3)$$



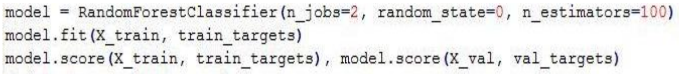
```
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier(random_state=42)
```

Fig. 7. Decision tree code snippet.

2) *Random Forest (RF):* The SKlearn module in Python is used to build the Random Forest technique as shown in Fig. 8, and 100 decision trees are employed to provide the final output for classification. The feature importance value for RF is calculated with the help of normalized importance of feature which are shown below in Eq. (4) and Eq. (5). The accuracy of the algorithm is coming 95%.

$$\mathrm{normfi}_i = fi_i / \sum_{j\in all\ features} fi_j \qquad (4)$$

$$\mathrm{RFfi}_i = \sum_j normfi_{ij} / \sum_{j\in all\ features, k\in all\ trees} normfi_{jk} \qquad (5)$$

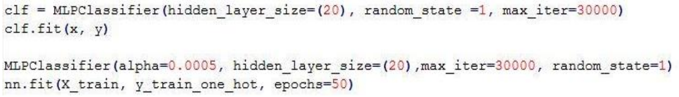


```
model = RandomForestClassifier(n_jobs=2, random_state=0, n_estimators=100)
model.fit(X_train, train_targets)
model.score(X_train, train_targets), model.score(X_val, val_targets)
```

Fig. 8. Random forest code snippet.

3) *Artificial Neural Network (ANN):* The input, hidden, and output layers of neurons are the three that are chosen in this procedure. There are 20 units in the hidden layer and a 0.1 starting random weight, for a total of 30000 iterations. The ANN code snippet is mentioned in the below Fig. 9. The accuracy of the algorithm is 90%. Eq. (6) shows the output layer calculation. Fig. 10 shows the following diagram illustrates the neural network used in this application. The neural network is too large to be plotted.

$$\mathrm{f(x)} = \sum_{i=1}^{m}(w_{ij}x_i) + bias \qquad (6)$$



```
clf = MLPClassifier(hidden_layer_size=(20), random_state =1, max_iter=30000)
clf.fit(x, y)

MLPClassifier(alpha=0.0005, hidden_layer_size=(20),max_iter=30000, random_state=1)
nn.fit(X_train, y_train_one_hot, epochs=50)
```

Fig. 9. ANN code snippet.

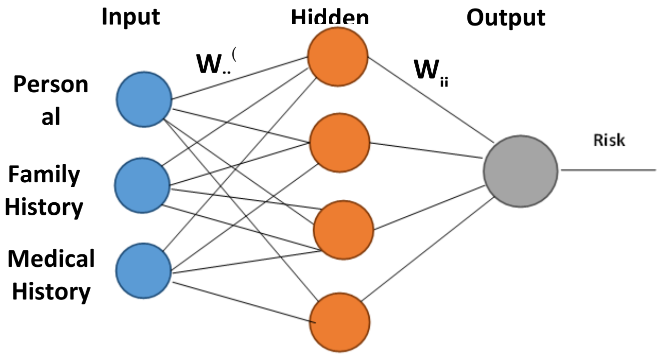
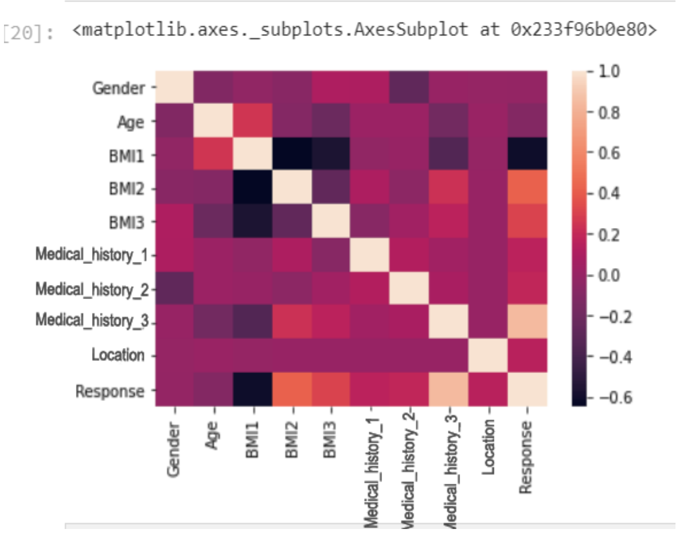


Fig. 10. A diagram of neural network for risk related application.

4) *Heat Map:* A heat map is a representation of two-dimensional data which shows the values having different intensities of colors. A simple heat map provides an immediate visual summary of information which is shown below in the Fig. 11. The heat map helps the viewer to understand and interpret the complex data sets. The color variation based on the different intensities provides understandable visual indications and also signifies the phenomenon of cluster or non-cluster over space.



```
[20]: <matplotlib.axes._subplots.AxesSubplot at 0x233f96b0e80>
```

Fig. 11. Heat Map of the insurance data.

IV. Results and Analysis

Here I have developed a web GIS-Dashboard using scripting languages like HTML, CSS, JavaScript as front end and PHP, python as back end. Geoserver is used here to publish different layers for displaying in this portal. Different base maps like Google map, Cartodb, Bhuvan map and Open street map were used here. In addition to its state boundary and district boundary of North East India were used here as overlay layers. Moreover, census data and three cycles of LULC data viz., LULC 1st cycle (2005-06), LULC 2nd cycle (2011-12) and LULC 3rd cycle (2015-16) are provided for analyzing and comparing trends over time. Different hazards like malaria (2019), covid-19 (2020-2021), MODIS air pollution data (2021) and industrial area (LULC 3rd cycle) were shown on this dashboard. Fig. 12 shows the UI of the dashboard. The default map that appears here is the Bhuvan map and NER district boundary on top of it. Different types of platforms are used for developing the web GIS-Dashboard which is as follows:

- Front End: HTML, CSS, JavaScript
- Back End : PHP, Python
- Web Server : XAMP, Geoserver
- Database : Postgres

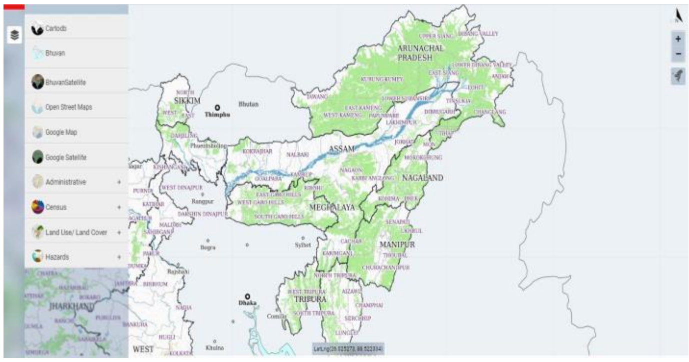


Fig. 12. Web portal designed using script language based environment.

Fig. 13 shows different base maps like Cartodb, Bhuvan, Open Street and Google map. In addition to it different overlay layers like state boundary, district boundary, census, LULC and different hazard layers are shown in this figure. The census data for Barpeta district is shown in Fig. 14, which includes total population, total work population, literacy, and the number of households.

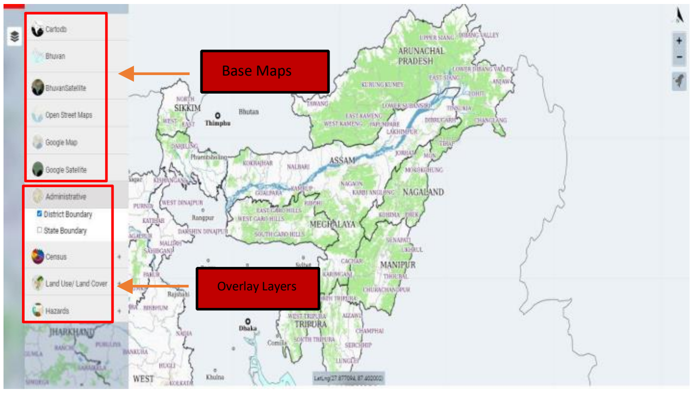


Fig. 13. Different base maps and overlay layers.

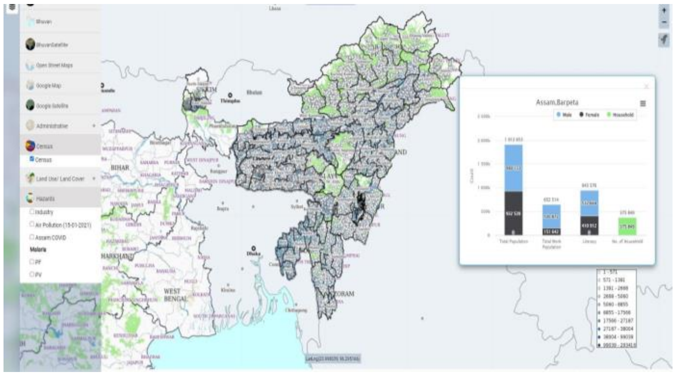


Fig. 14. Representation of census data for barpeta district.

Fig. 15 depicts the various LULC 1st cycle sectors for the Karbi Anglong district, including agriculture, built-up, forest, wasteland, and waterbodies. The analysis and comparison of all three LULC cycles in different sectors viz., agriculture, built up, waterbodies, snow, shifting cultivation, wastelands; forest for the Karbi Anglong area is shown in Fig. 16. In the Fig. 17, the risk level of different districts of Assam is depicted based on the insurance data and industrial data for the LULC 3rd cycle. The industrial area is shown in red marks for the NER region.

Fig. 18 displays the air pollution MODIS data for the NER region on January 15, 2021. Red indicates a high level of pollution, whereas blue indicates a low level of pollution. The figure shows various risk levels for insurance company applicants broken down by district. The following Figure 19 shows the corona virus count for the Assam district from January 2020 to October 2021. The graph shows the varied risk levels for insurance applicants in various districts of Assam.

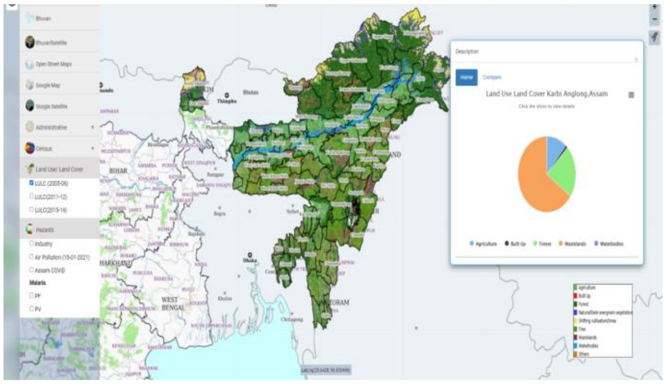


Fig. 15. Representation of LULC 1st cycle data for karbi anglong district.

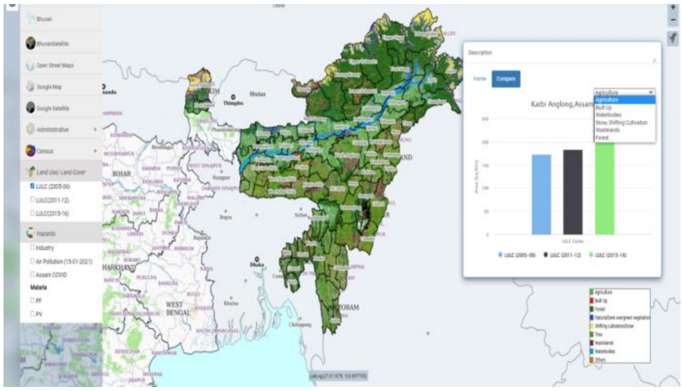
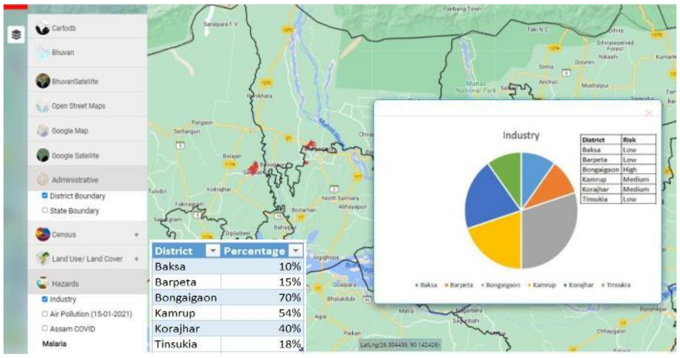


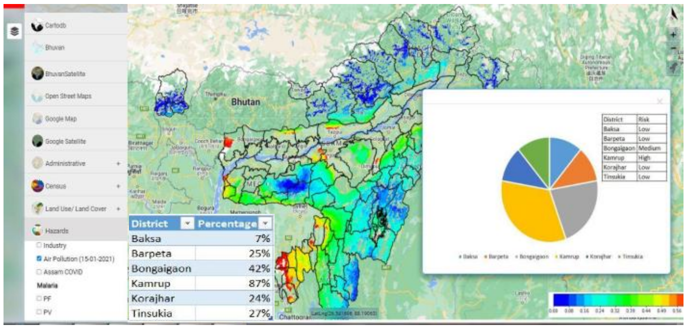
Fig. 16. Analysis and comparison of the LULC 3 cycle in several sectors for the karbi anglong district.



District	Risk
Baksa	Low
Barpeta	Low
Bongaigaon	High
Kamrup	Medium
Korajhar	Medium
Tinsukia	Low

District	Percentage
Baksa	10%
Barpeta	15%
Bongaigaon	70%
Kamrup	54%
Korajhar	40%
Tinsukia	18%

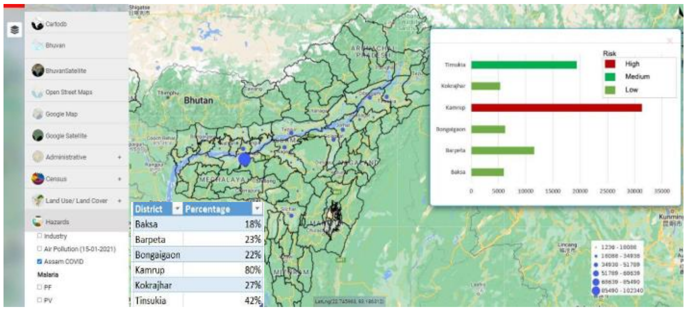
Fig. 17. Red marks on the map represent the industrial area for Assam district with risk level in various districts of Assam.



District	Percentage
Baksa	7%
Barpeta	25%
Bongaigaon	42%
Kamrup	87%
Korajhar	24%
Tinsukia	27%

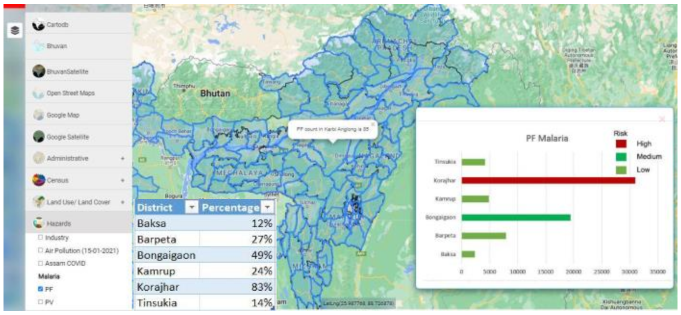
District	Risk
Baksa	Low
Barpeta	Low
Bongaigaon	Medium
Kamrup	High
Korajhar	Low
Tinsukia	Low

Fig. 18. MODIS data of NER air pollution for 15 January 2021 with risk level in various districts of Assam.



District	Percentage
Baksa	18%
Barpeta	23%
Bongaigaon	22%
Kamrup	80%
Kokrajhar	27%
Tinsukia	42%

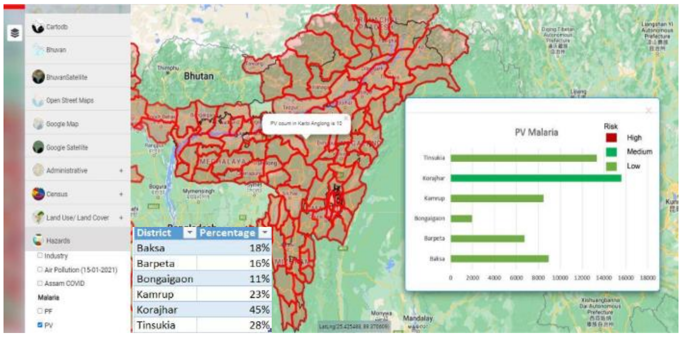
Fig. 19. Corona virus count from January 2020 to October 2021 with the risk level in various districts of Assam.



District	Percentage
Baksa	12%
Barpeta	27%
Bongaigaon	49%
Kamrup	24%
Korajhar	83%
Tinsukia	14%

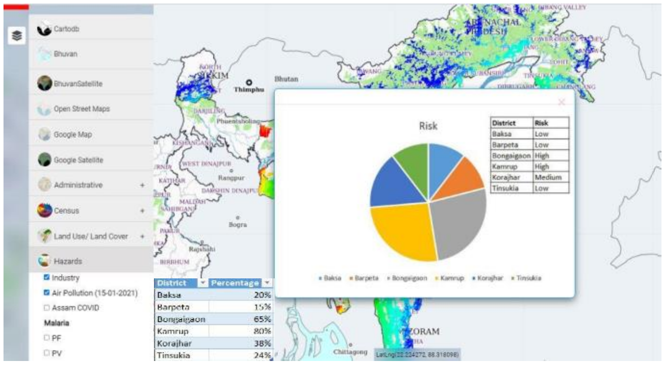
Fig. 20. Analysis of malaria- PF count in the infected area of India with different risk levels.

Fig. 20 displays the malaria-PF virus counts for various districts in India for the year 2019, and below Fig. 21 displays the malaria-PV virus numbers for various districts for the same year, with different risk levels for insurance company applicants for various districts of Assam. Fig. 22 shows the risk of the applicants in different districts of Assam based on two factors Industrial area and Air pollution. The result is depicted as pie chart and table as shown in the image. The risk of the applicants is based on three factors Industry, Air pollution and Covid for different districts of Assam as shown in Fig. 23. Fig. 24 shows the total risks of applicants in the insurance dataset based on three categories i.e., High, Medium and Low. The main insight and opinion offered by this research is that the integration of GIS and Machine Learning presents a promising and transformative approach for improving risk prediction in the Health Insurance Sector, providing valuable insights for industry stakeholders and policymakers.



District	Percentage
Baksa	18%
Barpeta	16%
Bongaigaon	11%
Kamrup	23%
Korajhar	45%
Tinsukia	28%

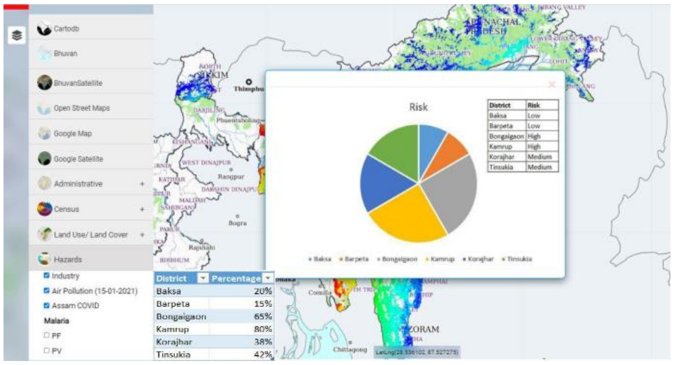
Fig. 21. Analysis of malaria- PV count in the infected area of India with different risk.



District	Percentage
Baksa	20%
Barpeta	15%
Bongaigaon	65%
Kamrup	80%
Korajhar	38%
Tinsukia	24%

District	Risk
Baksa	Low
Barpeta	Low
Bongaigaon	High
Kamrup	High
Korajhar	Medium
Tinsukia	Low

Fig. 22. Total risk of the applicants is calculated for different districts considering two factors industry and air pollution.



District	Percentage
Baksa	20%
Barpeta	15%
Bongaigaon	65%
Kamrup	80%
Korajhar	38%
Tinsukia	42%

District	Risk
Baksa	Low
Barpeta	Low
Bongaigaon	High
Kamrup	High
Korajhar	Medium
Tinsukia	Medium

Fig. 23. Total risk of the applicants is calculated for different districts considering three factors industry, air pollution and covid.

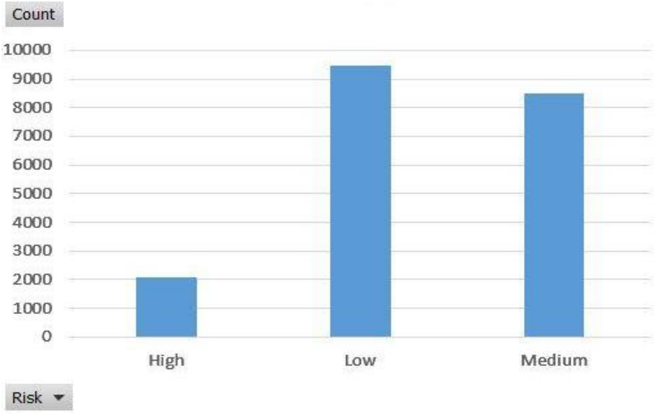


Fig. 24. Total risk of the applicants in 3 categories.

V. Conclusion

A GIS software model that makes use of online GIS technologies for analysis and visualization purposes which have been developed in this paper. The insurance provider might therefore be in a position to conduct a more thorough investigation and come to better judgments for its customers and business. Additionally, the corporate environment will specifically benefit from this research endeavour. Data visualization and analysis are becoming more and more common among enterprises all around the world. In this paper, several factors are incorporated in QGIS which provides risk analysis in better interpretable form. When compared to conventional methods, predictive modeling using GIS technologies can have a substantial impact on how business is performed in the life insurance market. Previously, lengthy and difficult actuarial calculations were used to evaluate risk for life underwriting. With the use of a web GIS map, the task may now be accomplished more quickly and with better results. As a result, it would benefit the company by enabling quicker customer service, thereby boosting satisfaction and loyalty.

The future research development is being planned in the following ways:

- Calculating premium for the customers in a specific geographic location based on the prediction.
- Collecting data in large quantities in order to increase the accuracy and usability of the model in real life situations.

References

[1] I. Franch-Pardo, B. M. Napoletano, F. Rosete-Verges, and L. Billa, “Spatial analysis and GIS in the study of COVID-19. A review,” Science of the Total Environment, vol. 739, 2020.

[2] M. H. Rahman, N. M. Zafri, Fajle Rabbi Ashik, Md Waliullah, A. Khan, “Identification of risk factors contributing to COVID-19 incidence rates inBangladesh: A GIS-based spatial modeling approach,” Heliyon, vol. 7, pp. e06260, 2021.

[3] N. Maged, K. Boulos, E.M. Geraghty, “Geographical tracking and mapping of coronavirus disease COVID 19/severe acute respiratory syndrome coronavirus 2 (SARS CoV 2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fightagainst outbreaks and epidemics,” International Journal ofHealth Geographics, vol. 19, 2020.

[4] N. Boodhun, Jayabalan, “M. Risk prediction in life insurance industry using supervised learningalgorithms,” *Complex & Intelligent Systems*, vol. 4, pp. 145-154, 2018.

[5] H. Angier, S. Likumahuwa, S. Finnegan, T. Vakarcs, C. Nelson, A. Bazemore, M. Carrozza, J. E. DeVoe, Using Geographic Information Systems (GIS) to Identify Communities in Need of Health Insurance Outreach: An OCHIN Practice-based ResearchNetwork (PBRN) Report. *JABFM* 27, 804-810, 2014.

[6] S. Saran, P. Singh, V. Kumar, P. Chauhan, “Review of Geospatial Technology for Infectious Disease Surveillance: Use Case on COVID-19,” *Journal of the Indian Society of Remote Sensing*, vol. 48, pp. 1121–1138, 2010.

[7] Changing face of the Insurance Industry. Available online: https://www.infosys.com/industries/insurance/whitepapers/ Documents/changing-face-insurance-industry,pdf (accessed on 09 November 2023).

[8] K. M. Al Kindi, A. Alkharusi, D. Alshukaili, N. A. Nasiri, T. A. Awadhi, Y. Charabi, A. M. E. Kenawy, “Spatiotemporal Assessment of COVID 19 Spread over Oman Using GIS Techniques,” *Earth Systems and Environment*, vol. 4, pp. 797-811, 2020.

[9] A. Das, A.R. Anvikar, L. J. Cator, R. C. Dhiman, A. Eapen, N. Mishra, B. N. Nagpal, N. Nanda, K. Raghavendra, A. F. Read, S. K. Sharma, O. P. Singh, V. Singh, P. Sinnis, H. C. Srivastava, S. A. Sullivan, P. L. Sutton, M. B. Thomas, J. M. Carlton, N. Valecha, “Malaria in India: The Center for the Study of Complex Malaria in India,” *Acta Trop.*, vol. 121, pp. 267-273, 2012.

[10] M. R. Rahman, A.H.M.H. Islam, M. N. Islam, “Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID 19) pandemic in Bangladesh towards vulnerability zoning and management approaches,” *Modeling Earth Systems and Environment*, vol. 7, pp. 2059-2087, 2021.

[11] B.J. Hutagaol, T. Mauritsius, “Risk level prediction of life insurance applicant using machine learning,” International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, pp. 2213-2220, 2020.

[12] R. Jain, J.A. Alzubi, N. Jain, P. Joshi, “Assessing risk in life insurance using ensemble learning,” Journal of Intelligent & Fuzzy Systems, vol. 37, pp. 2969-2980, 2019.

[13] R. Biddle, S. Liu, P. Tilocca, G. Xu, “Automated underwriting in life insurance: Predictions and optimisation.&quot; Databases Theory and Applications,” *In the proceedings of 29th Australasian Database Conference (ADC)*, Gold Coast, QLD, Australia, 2018.

[14] W. F. Mustika, H. Murfi, Y. Widyaningsih, “Analysis Accuracy of XGBoost Model for Multiclass Classification - A Case Study of Applicant Level Risk Prediction for Life Insurance,” *In Proceedings of the 5th International Conference on Science in Information Technology* (ICSITech), Yogyakarta, Indonesia, pp. , 2019.

[15] S. Dwivedi, A. Mishra, A. Gupta, “Risk Prediction Assessment in Life Insurance Company through Dimensionality Reduction Method,” International Journal of Scientific & Technology Research, vol. 9, pp. 1528-1532, 2020.

[16] S. Roy, S.K. Ojah, N. Nishant, P.P. Singh, D. Chutia, “Spatio-temporal Analysis of Flood Hazard Zonation in Assam,” In: Gupta, D., Goswami, R.S., Banerjee, S., Tanveer, M., Pachori, R.B. (eds.): LNEE, vol. 888, Springer, Singapore, pp. 521-531, 2022.

[17] P. P. Singh, B. Das, U. Poddar, D. R. Choudhury, and S. Prasad, “Classification of diabetic’s patient data using machine learning techniques,” In: Perez, G., Tiwari, S., Trivedi, M., Mishra, K. (eds.): Ambient Communications and Computer Systems. Advances in Intelligent Systems and Computing, vol. 696, Springer Singapore, pp. 427-436, 2017.

Enhancing Underwater Object Recognition Through the Synergy of Transformer and Feature Enhancement Techniques

Hoanh Nguyen*, Tuan Anh Nguyen
Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

Abstract—**Underwater object recognition presents a unique set of challenges due to the complex and dynamic characteristics of marine environments. This paper introduces a novel, multi-layered architecture that leverages the capabilities of Swin Transformer modules to process segmented image patches derived from aquatic scenes. A key component of our approach is the integration of the Feature Alignment Module (FAM), which is designed to address the complexities of underwater object recognition by enabling the model to selectively emphasize essential features. It combines multi-level features from various network stages, thereby enhancing the depth and scope of feature representation. Furthermore, this paper incorporates multiple detection heads, each embedded with the innovative ACmix module. This module offers an integrated fusion of convolution and self-attention mechanisms, refining detection precision. With the combined strengths of the Swin Transformer, FAM, and ACmix module, the proposed method achieves significant improvements in underwater object detection. To demonstrate the robustness and effectiveness of the proposed method, we conducted experiments on the UTDAC2020 dataset, highlighting its potential and contributions to the field.**

Keywords*—*Underwater object recognition; swin transformer; self-attention; feature alignment

I. INTRODUCTION

Underwater object recognition is a specialized domain within computer vision and robotics that focuses on identifying and locating objects within aquatic environments. The complexities associated with this field are manifold, given the unique challenges posed by underwater conditions. These include limited visibility due to turbidity, light refraction and attenuation, and the dynamic nature of the aquatic medium with constantly moving particles and organisms. Detecting objects in such environments is crucial for a variety of applications, ranging from marine biology research, underwater archaeology, to defense and surveillance. Advanced techniques and algorithms in this area not only aim to improve the accuracy of detection but also enhance the real-time processing capabilities, making it possible for autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs) to perform intricate tasks with minimal human intervention. Traditional underwater object detection is usually based on handcrafted features of images for detecting objects [1], [2], [3]. In these conventional methods, detection was often based on basic image processing techniques. Specifically, thresholding, contour detection, and basic filter operations were commonly employed to differentiate objects from the surrounding environment. While these methods had their merits, especially in low-visibility conditions, they often struggled with false positives and lacked the precision needed for intricate tasks. Furthermore, these approaches were highly dependent on manual calibration and expert interpretation, making them labor-intensive.

In recent years, deep learning has revolutionized the field of object detection, ushering in a new era of accuracy and efficiency. These methods leverage complex neural network architectures, particularly Convolutional Neural Networks (CNNs), to automatically learn hierarchies of features from raw pixel data, eliminating the need for handcrafted feature extraction. Advanced architectures such as Faster R-CNN [4], YOLO [5], [6], SSD [7], R-FCN [8], Mask R-CNN [9] have emerged as frontrunners, offering real-time detection capabilities with impressive precision. These models have been applied in various vision applications such as depth estimation [10], intrusion detection [11], [12], vehicle license detection [13], and face mask detection [14]. With the success of deep learning-based object detection models, researchers have begun to apply deep learning to underwater object detection [15-24]. Although these methods have achieved certain successes, they encounter a number of issues. Firstly, all objects, regardless of their ambiguity, are subjected to the same supervisory signal. As a result, the classification scores obtained using simple cross-entropy loss don't accurately represent the ambiguity of the objects. This leads to misleadingly overconfident predictions. Secondly, these methods struggle with objects that are vague due to blurred boundaries or colors similar to their background. This similarity makes it challenging for the methods to distinguish such objects from their surroundings effectively.

Recognizing these limitations, our study aims to overcome these specific challenges. We propose an innovative approach for underwater object detection leveraging a multi-layered framework powered by the Swin Transformer. In the model, images pass through a patch partitioning process, segmenting them into smaller patches. These patches are then processed through several Swin Transformer layers. After each transformation, techniques like concatenation, upsampling, and convolution are utilized to enrich the feature maps. These enhanced feature maps pass through the FAM to amplify feature representation, ensuring precise object detection in complex underwater scenarios. Following the FAM, the framework integrates multiple detection heads, ensuring reliable localization of objects in underwater imagery. By addressing the core issues of overconfidence in predictions and the struggle with vague object boundaries, our method seeks to provide a more robust and accurate solution for object detection.

The remainder of this paper is organized as follows: Section II provides an in-depth review of related work. Section III details our proposed methodology. In Section IV, we present a thorough analysis of our experiments. Finally, Section V concludes the paper with a summary of our findings and implications for future research.

II. Related Work

Underwater object detection, a crucial technology enabling AUVs to execute various tasks beneath the surface, has garnered significant interest globally among researchers. In [15], the authors introduced a method that utilizes a region proposal network from Faster R-CNN to enhance underwater object detection and recognition speed. This approach achieves quicker detection by employing convolutional networks to produce superior object candidates and integrating these networks with the primary detection systems. Chen et al. [16] proposed the Sample-WeIghted hyPEr Network (SWIPENET) and the Curriculum Multi-Class Adaboost (CMA) training paradigm to address challenges in underwater object detection, specifically blurry images with noise and small object detection. SWIPENET uses Hyper Feature Maps for enhanced resolution and detection of small objects, while its sample-weighted detection loss function emphasizes learning from high weight samples and disregarding low weight ones. Wei et al. [17] addressed challenges in underwater image target detection, particularly blur caused by water particles, by integrating squeeze and excitation modules into the YOLOv3 model after its deep convolution layers, enhancing semantic information. Zeng et al. [18] introduced the Faster R-CNN-AON network by integrating an adversarial occlusion network (AON) with the standard Faster R-CNN detection algorithm. The AON competes with the Faster R-CNN, teaching it to obscure targets, which in turn enhances the robustness of underwater seafood detection and prevents overfitting of the detection network.

In another approach, Lingyu et al. [19] adapted the YOLOv4 neural network for underwater target recognition by substituting its upsampling module with a deconvolution module and integrating depthwise separable convolution. Cao et al. [20] addressed underwater dynamic target tracking by developing a deep learning-based detection algorithm that uses the YOLO v3 network to identify targets in multibeam forward-looking sonar images and determine their positions. Huang et al. [21] introduced three specialized data augmentation techniques to address the scarcity of labeled samples in underwater environments: the inverse process of underwater image restoration for creating varied marine turbulence scenarios, perspective transformation to simulate different camera viewpoints, and illumination synthesis for replicating uneven lighting conditions underwater. In study [22], an innovative underwater salient object detection method that integrates both 2D and 3D visual features was introduced. This approach combines color and intensity (2D features) with 3D depth features, enhanced by a region-specific method that separately extracts these features in artificial and natural light regions, leading to more comprehensive and accurate detection results in three-dimensional underwater environments. Lin et al. [23] focused on augmentation policies designed to simulate overlapping, occluded, and blurred objects, constructing a model that achieves enhanced generalization. They introduce RoIMix, an innovative augmentation method that blends proposals from multiple images, unlike previous methods that operate on single images, thereby creating more complex and varied training data to improve model performance. Recently, Song et al. [24] introduced a two-stage underwater detector called boosting R-CNN, which features a novel region proposal network named RetinaRPN for high-quality proposals and models object prior probability through objectness and IoU prediction.

III. Proposed Model

A. Overview Pipeline

Fig. 1 illustrates the overall structure of our method for underwater object detection. The proposed method employs a multi-layered architecture that exploits the power of Swin Transformer modules. The input underwater image is first processed through a patch partition module, which segments the image into manageable patches. These patches are then sequentially passed through four layers of Swin Transformer modules. Specifically, Layers 1 and 2 involve two repetitions of the Swin Transformer module, Layer 3 contains six repetitions, while Layer 4 processes the patches twice through the Swin Transformer module. After the transformation process in each layer, specific operations including concatenation, upsampling, and convolution are performed to refine the feature maps. These refined feature maps are then passed through the Feature Alignment Module (FAM) to further enhance the feature representation, ensuring accurate object detection in the complex underwater environment. Following the FAM, the architecture incorporates multiple detection heads, which are responsible for the final object detection, ensuring robust identification of objects present in the underwater image. The details of each module are explained in the following subsections.

B. Swin-Transformer Backbone

- Transformers, initially introduced by Vaswani et al. in 2017 [25], are a type of neural network architecture primarily designed for handling sequence-to-sequence tasks in the field of natural language processing (NLP). They make use of attention mechanisms, notably self-attention, to weigh the significance of different parts of the input data. While Transformers have achieved remarkable success in NLP, their direct application to the vision domain presents challenges. One major reason is that unlike textual data which is inherently sequential, images are spatially structured with local patterns and hierarchies. Processing an image as a flat sequence of pixels loses this spatial coherence. Additionally, due to the high-dimensionality of images, Transformers can be computationally expensive and memory-intensive. To address these challenges, various adaptations, such as Vision Transformers (ViTs) [26] which divide images into fixed-size patches and then linearly embed them, have been proposed to better suit the unique characteristics of visual data. However, they sometimes lack fine-grained local feature extraction. Recently, Swin Transformer [27] has emerged as a novel and powerful architecture that brings together the strengths of both classic CNNs and Transformers, and in some contexts, it has outperformed both. While CNNs have traditionally been strong at capturing local features through their hierarchical design of convolutional layers, they often struggle with long-range dependencies and global context. Swin Transformer tackles the issues of both CNNs and Transformers by hierarchically partitioning the image into non-overlapping windows and applying shifted windows across layers. This approach allows it to capture both local features within each window and global context across the entire image. The combination of local window-based processing with the global contextual understanding provided by the Transformer structure makes Swin Transformer particularly effective at feature extraction, offering advantages over traditional CNNs and basic Transformers. Underwater images often exhibit a range of complexities, including varying light conditions, attenuation, backscatter, and color distortions. Traditional architectures, like CNNs, can sometimes struggle with these irregularities, especially when it comes to recognizing objects that may be obscured or distorted due to water turbidity. Swin Transformer, with its hierarchical partitioning and shifted windows, can capture both local details and global contexts effectively. The local window-based processing ensures fine-grained feature extraction, which is crucial for identifying subtle characteristics of underwater objects. Meanwhile, the global contextual understanding inherent in the Transformer structure helps in identifying objects even when they are partially obscured or when the surrounding environment is cluttered. Additionally, the self-attention mechanism of the Swin Transformer can focus on long-range dependencies, which is beneficial for analyzing the spatial relationships between various underwater elements. Based on the features analyzed above, we chose Swin Transformer as the backbone network to perform feature extraction.

- The architecture of Swin Transformer is depicted in Fig. 2. It takes an image of dimensions *w*×*h*×*c* (where *c* denotes channels, *h* is height, and *w* is width) and partitions it into non-overlapping patches. These patches pass through a linear embedding transformation and patch merging operations, converting them into token representations suitable for processing by transformer blocks. As the image progresses through the layers of the architecture (Layer 1 to Layer 4), the spatial resolution decreases, while the embedding dimensions increase. Specifically, the resolutions get adjusted by a downsampling factor (*w*/4×*h*/4) to (*w*/32×*h*/32) from Layer 1 to Layer 4. Each stage contains a specific number of Swin Transformer blocks, denoted by multipliers (i.e., ×2, ×2, ×6, ×2).

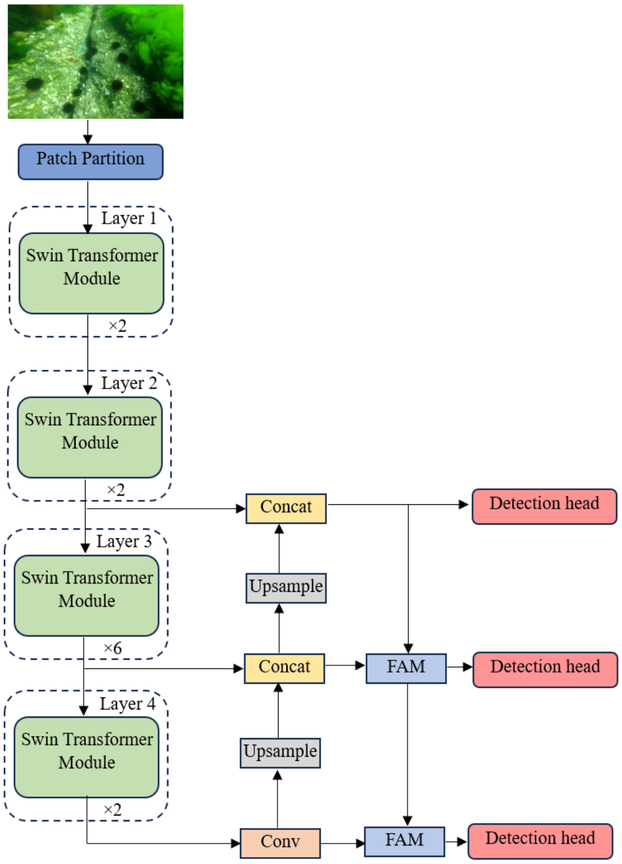


Fig. 1. Overview of the proposed model.

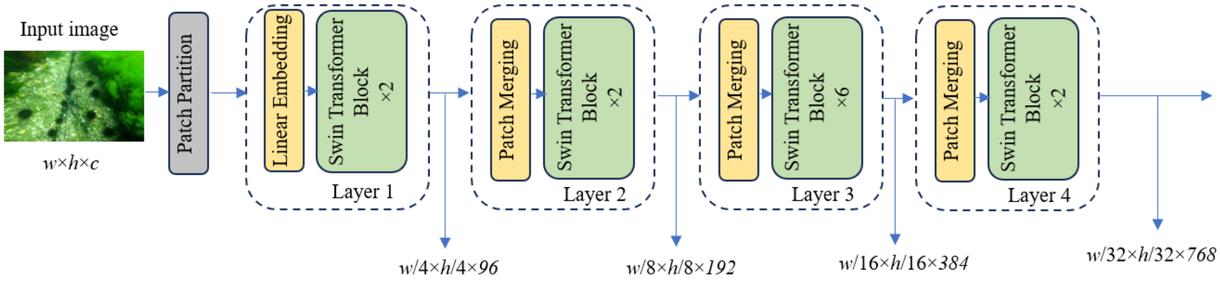


Fig. 2. Swin transformer architecture.

1) Patch partition block: Given an input image $I \in R^{w \times h \times c}$, the patch partition block divides this image into non-overlapping patches of size $p \times p \times c$. The total number of patches, N, produced by this partitioning is given by:

$$N = \left(\frac{w}{p}\right) \times \left(\frac{h}{p}\right) \qquad (1)$$

Each patch is then flattened to produce a vector of dimension $p^2 \times c$. Thus, after the patch partition block, the image representation transforms from I to a matrix P of dimensions $N \times (p^2 \times c)$. In essence:

$$I \in R^{W \times H \times C} \rightarrow P \in R^{N \times (p^2 \times c)} \qquad (2)$$

where, $P[i]$ represents the flattened vector for the i^{th} patch.

2) Linear embedding block: The linear embedding block applies a linear transformation to each of these flattened patches to project them into a specified embedding dimension D. This transformation can be represented by a matrix E of dimensions $(p^2 \times c) \times D$. Thus, the output L of the linear embedding block for each patch can be computed as:

$$L[i] = P[i] \times E \qquad (3)$$

where, $L[i]$ represents the embedded vector for the i^{th} patch.

Given this, the entire input-output relationship for the linear embedding block can be represented as:

$$P \in R^{N \times (p^2 \times c)} \rightarrow L \in R^{N \times D} \qquad (4)$$

where, P is the matrix of flattened patches, L is the embedding matrix.

3) Swin transformer block: The structure of two successive Swin Transformer blocks is depicted in Fig. 3. Each block consists of a sequence of operations: Layer Normalization (LN), Window-based Multi-head Self Attention (W-MSA) or Shifted Window-based Multi-head Self Attention (SW-MSA), and a Multi-Layer Perceptron (MLP) head. The LN standardizes the activations by calculating the mean and variance of the input image patch, thus stabilizing the training process. The W-MSA operates on a set of query (Q), key (K), and value (V) vectors. It matches the query to a set of key-value pairs to produce an output. This matching is achieved by computing the dot product between the query vector and every key vector. Subsequently, a softmax function is employed to scale these dot products, transforming them into weights denoted as k. The process is calculated as follows:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (5)$$

where, Q represents the Query matrix, K represents the Key matrix, V stands for the Value matrix, and d_k is the dimension of the keys. The divisions by $\sqrt{d_k}$ functions as a scaling factor, ensuring stability in the gradients during the training phase.

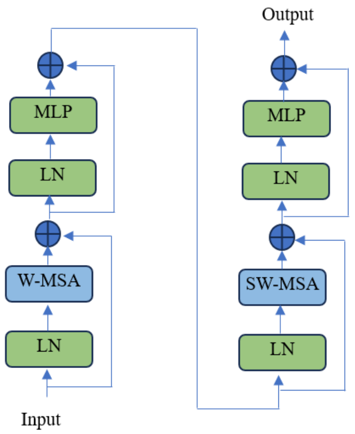


Fig. 3. The structure of two successive swin transformer blocks.

The locality of W-MSA might raise concerns about its ability to capture global context. To mitigate this, Swin Transformer employs multiple blocks of W-MSA and integrates a "shifting" strategy in subsequent blocks (SW-MSA), ensuring that tokens in one window in a certain block can interact with tokens in neighboring windows in the next block.

4) Patch merging block: This block is used to reduce the spatial dimensions of the input while augmenting the feature dimensions. Conceptually, this block aggregates neighboring patches from the previous layer and fuses them to form a larger patch. For instance, four adjacent patches of size p × p are merged to create a single patch of size 2p × 2p. This merging process typically employs a simple linear transformation. As a result, the spatial resolution of the feature map is halved in both height and width dimensions, but the depth or the number of channels is doubled. The purpose of this operation is twofold: firstly, it progressively reduces the computational requirements for subsequent layers, and secondly, it increases the receptive field, enabling the model to capture more global and abstract features as information flows deeper into the transformer.

C. *Feature Attention Mechanism*

In hierarchical models such as Swin Transformers and CNNs, lower-level features often capture fine-grained details, textures, and simple patterns. Meanwhile, higher-level features encompass more abstract, complex, and semantically rich information about objects, enabling the model to recognize more intricate and high-level attributes. By combining features from different layers, the model is equipped with a comprehensive and multi-scale representation of the input image. In addition, the underwater visuals are typically characterized by low-light conditions, varied light absorption and scattering, and blurry images due to particulate matter suspended in the water, which can result in a significant degradation of image quality and object distinguishability. This paper proposes a feature attention mechanism (FAM) to precisely address these challenges by enabling the model to selectively focus on important features and effectively integrate multi-level features from different stages of the network, enhancing the representative power of deep features, especially in the challenging context of underwater object detection. The architecture of the FAM is illustrated in Fig. 4, which consists of two branches: the first branch directly processes the lower-level feature (F_1) through a batch normalization layer, ensuring the features are normalized and thereby enhancing the model's stability and convergence during training. Simultaneously, the second branch takes the higher-level feature (F_2) through a sophisticated pathway comprising a coordinate attention (CA) block, followed by a convolution layer, a max-pooling layer, and a batch normalization layer. The CA block [28] is notable for its capacity to encode both channel relationships and long-range dependencies with precise positional information, implemented through two crucial steps: coordinate information embedding and coordinate attention generation. Once the individual pathways of the two branches have processed the features, their outputs are aggregated by summation and then fed to a ReLU activation layer, which ensures the generation of a robust and hierarchically rich feature representation, designed to significantly enhance the underwater object detection capabilities of the system. The output of the FAM mechanism can be calculated as follows:

$$F_1' = BN(F_1) \tag{6}$$

$$F_2' = BN(MAXPOOL(CONV(CA(F_2)))) \tag{7}$$

$$F_{output} = ReLU(F_1' + F_2') \tag{8}$$

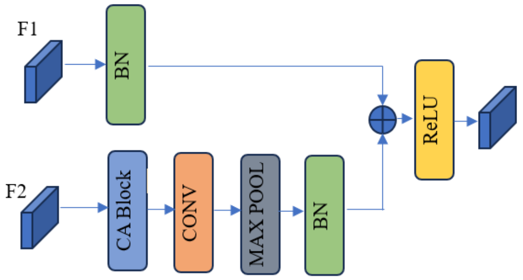


Fig. 4. Feature attention mechanism.

D. Detection Head with Combination of Convolution and Self- Attention

Pan et al. [29] highlighted the connection between convolution and self-attention mechanisms by highlighting computational similarities in both methods. Consequently, they designed a hybrid model, ACmix, which adeptly integrates the advantages of both self-attention and convolution, while maintaining minimal computational overhead relative to pure convolution or self-attention models. Given the potentially robust link between convolution and self-attention, the ACmix module is utilized in this paper to integrate the convolution and self-attention mechanisms. Fig. 5 illustrates the structure of the ACmix module. The module channels the data through a series of three 1×1 convolutional layers. These layers serve to capture local features and correlations in the data. Concurrently, the input is also processed through a self-attention mechanism equipped with a position encoder. This mechanism ensures the model can recognize and weigh global dependencies in the data effectively. Subsequent to their independent processing, the outputs from the convolutional and self-attention pathways pass through a 'Shift Operation' and are concatenated. This combined representation exploits the strengths of both paradigms, ensuring a comprehensive understanding of the data's local and global patterns. Finally, the concatenated output is summed to produce the final output. The output of each ACmix module is input into a YOLO detector head for location and classification.

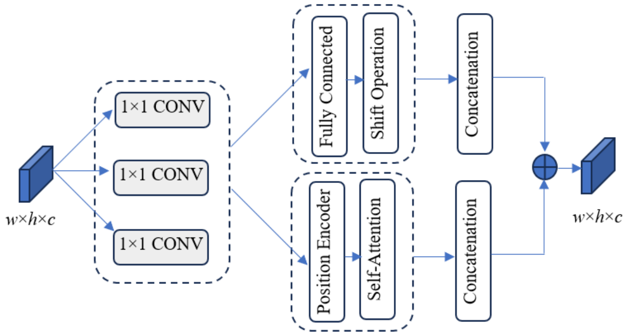


Fig. 5. The structure of ACmix.

IV. Experiments and Results

A. Dataset

The experiments utilize the UTDAC2020 dataset, which originates from the Underwater Target Detection Algorithm Competition in 2020 [18]. This comprehensive dataset comprises 5,168 training images and 1,293 validation images, focusing primarily on four specific marine species: echinus, holothurian, starfish, and scallop. The unique attribute of this dataset lies in its variety of resolutions, with images spanning four distinct sizes: 3840×2160, 1920×1080, 720×405, and 586×480. This dataset serves as an important resource for understanding and advancing underwater image analysis and target detection. For evaluation and comparison purposes, the standard COCO-style evaluation metric is employed.

B. Experimental Settings

The proposed model was implemented using the PyTorch deep learning framework and programmed in Python. All experiments were carried out on machines equipped with an NVIDIA RTX 4080 GPU. The backbone of our architecture is the base version of the Swin Transformer, which was pre-trained on the ImageNet-1K dataset and has an embedding dimension of C = 128. We chose this version because of its balance between computational efficiency, model size, and accuracy. The model was fine-tuned for 15 epochs, using a batch size of 2. For optimization, the AdamW optimizer [30] was employed, starting with a learning rate of 0.0002. This rate was adaptively adjusted based on the training progress, and a weight decay of 0.05 was implemented. Our data augmentation strategies included a variety of techniques such as random resizing, combined random resizing and cropping, as well as horizontal and vertical random flipping. For a comprehensive overview of the hyperparameters employed in the comparative models, (see Table I).

TABLE I. HYPERPARAMETERS OF ALL MODELS

Model	Initial learning rate	Regularizer	Optimizer	Batch size	Number of Epochs
Our model	0.0002	Weight decay of 0.05	AdamW	2	15
Deformable DETR [28]	0.00001	Weight decay of 0.0001	AdamW	2	40
RetinaNet [29]	0.01	L2	SGD with Momentum	10	20
Faster R-CNN with FPN [30]	0.02	Weight decay of 0.0001	SGD with Momentum	2	12
DetectoRS [31]	0.02	Weight decay of 0.0001	SGD with Momentum	2	20
FCOS [32]	0.01	Weight decay of 0.0001	SGD with Momentum	16	20
CenterNet [33]	0.0002	Weight decay of 0.0001	Adam	6	25

C. Comparison with Other Methods

The comparison results on the UTDAC2020 dataset are shown in Table II. Our underwater object detection model based on the Swin Transformer architecture obtains a significant improvement in performance when compared to other state-of-the-art models on the UTDAC2020 dataset. In terms of Average Precision (AP), the proposed model achieved a score of 51.6, which is notably higher than other models utilizing the ResNet50 backbone, such as Deformable DETR [31], RetinaNet [32], and Faster R-CNN with FPN [33], DetectoRS [34], FCOS [35], and especially CenterNet [36] which used ResNet18. Additionally, in the specific AP metrics (AP_{50}, AP_{75}), our Swin Transformer-based model also outperforms the competition, indicating a robustness in detecting objects at different Intersection over Union (IoU) thresholds. Remarkably, there is a significant jump in AP_{75} to 57.5, suggesting that the model is efficient at achieving a tighter fit around the detected objects. When analyzing the performance based on object size (AP_S, AP_M, AP_L), the proposed model consistently delivers superior results. The model's weakest performance is in AP_S at 23.2, which, while comparable to some models like Deformable DETR and DetectoRS, demonstrates that there might be challenges in detecting smaller underwater objects. Nonetheless, the model's AP_M and AP_L scores of 44.6 and 57.9, respectively, emphasize its efficiency in medium to large object detection. In summary, leveraging the Swin Transformer architecture and feature attention mechanisms has enhanced the efficacy of the proposed model in the challenging domain of underwater object detection.

Fig. 6 shows qualitative results of our model on the UTDAC2020 dataset. We can see a notable performance of the proposed model across diverse underwater scenarios. The detection results are evident across a range of images, from those where marine life is interspersed among a sea of green to those with rocky terrains. Even in images with dense clusters of organisms or potential overlapping instances, the model is efficient in differentiating between individual entities, avoiding much of the occlusion-related errors that often plague underwater detection tasks. Furthermore, the model's performance is evident in various lighting conditions and water turbidities, emphasizing its robustness.

D. Importance of Feature Attention Mechanism

We also conducted experiments to evaluate the impact of the FAM. Fig. 7 shows comparing the performance of the model with and without the FAM. When FAM is implemented, there is a noticeable improvement in all metrics. Specifically, the AP increases from 50.1% to 51.6%, indicating a more accurate model overall. This improvement is more pronounced in AP_{50} (from 82.2% to 85.1%), which measures precision at 50% IoU threshold, suggesting that FAM particularly enhances the model's ability to detect objects with a moderate overlap with the ground truth. The increase in AP_{75}, from 54.2% to 57.5%, also highlights better performance at a stricter IoU threshold, implying enhanced precision for more accurately localized predictions. The improvements in AP_S, AP_M, and AP_L are also noteworthy. The model with FAM achieves better results across these size-based categories, with the most significant jump observed in the medium-sized object category, from 40.3% to 44.6%. This suggests that FAM effectively enhances the model's capability to recognize and detect objects of various sizes, especially in challenging underwater environments where visibility and image quality are often compromised. Overall, the integration of FAM into the model clearly leads to better performance in object detection, making it a valuable addition to the model's architecture, particularly for tasks in complex and visually challenging environments like underwater object detection.

TABLE II. COMPARISON RESULTS ON THE UTDAC2020 DATASET

Model	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Deformable DETR [31]	ResNet50	46.6	84.1	47.0	24.1	42.4	51.9
RetinaNet [32]	ResNet50	43.9	80.4	42.9	18.1	38.2	50.1
Faster R-CNN with FPN [33]	ResNet50	44.5	80.9	44.1	20.0	39.0	50.8
DetectoRS [34]	ResNet50	47.6	82.8	49.9	23.1	41.8	54.2
FCOS [35]	ResNet50	43.9	81.1	43.0	19.9	38.2	50.4
CenterNet [36]	ResNet18	31.3	61.1	27.6	11.9	32.5	33.4
Our model	Swin Transformer	51.6	85.1	57.5	23.2	44.6	57.9

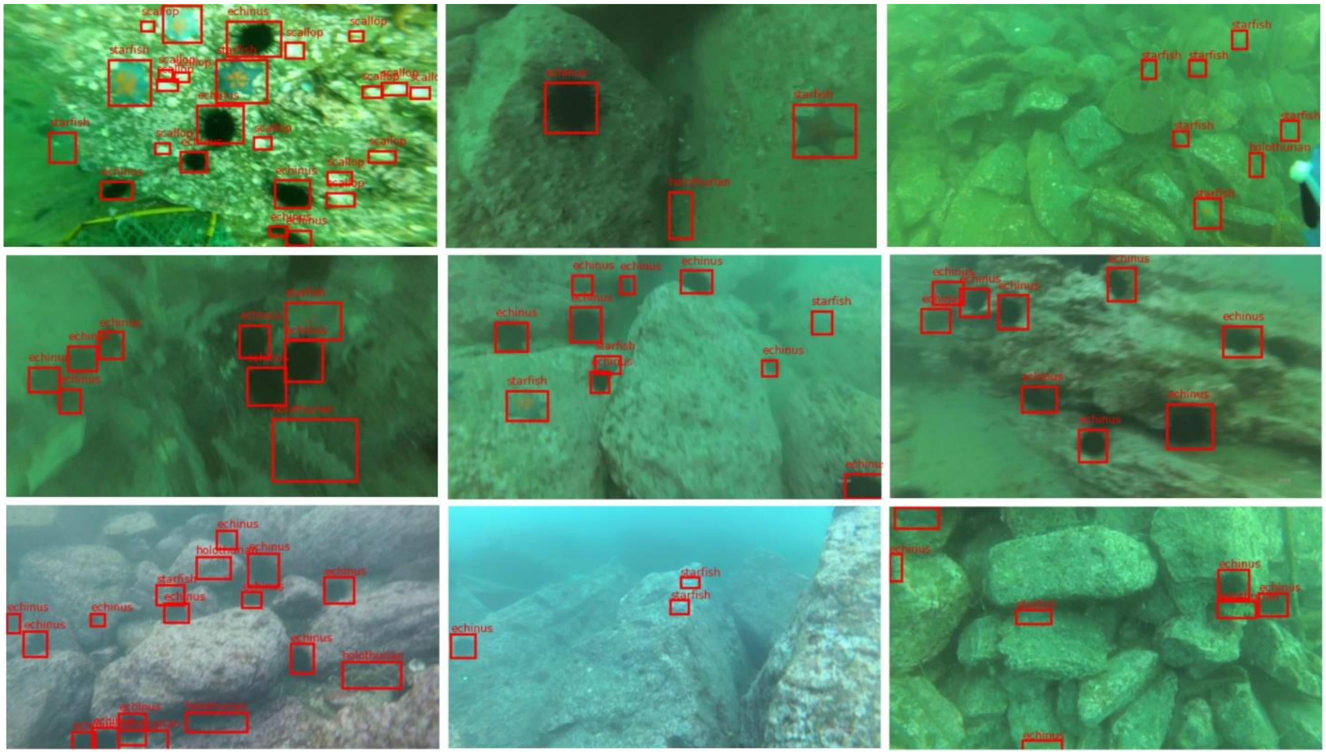


Fig. 6. Qualitative results on the UTDAC2020 dataset.

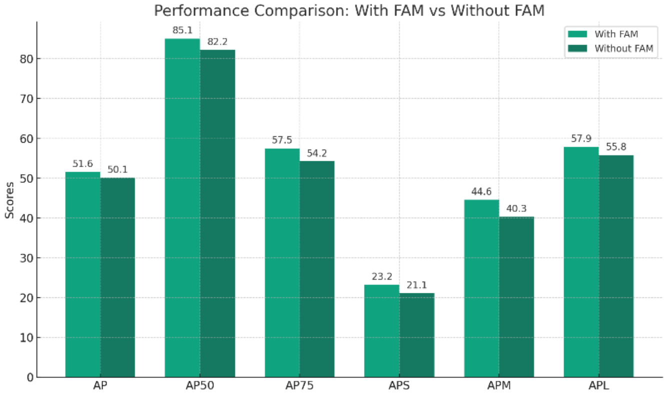


Fig. 7. Comparing the performance of the model with and without the FAM.

V. Conclusions

In this research, we addressed the challenges associated with underwater object detection by introducing a novel multi-layered architectural approach that effectively exploits the capabilities of Swin Transformer. Our method provides a structured approach to process segmented image patches derived from underwater scenes, ensuring accurate and efficient object detection. A significant contribution of our research is the Feature Alignment Module (FAM), specifically designed to address the complexities of marine environments. By focusing on essential features and integrating multi-level features across various network stages, the FAM substantially elevates the depth and precision of feature representation. Moreover, the incorporation of several detection heads, coupled with the ACmix module, represents a transformative approach to enhancing detection accuracy. The results on the UTDAC2020 dataset emphasize not only the efficacy of our proposed method but also its potential as a benchmark solution in the field of underwater object detection. In future, we envision further refining our model by integrating more advanced attention mechanisms and exploring its applicability in other complex environmental scenarios.

References

[1] Shi, X. U. X., and J. L. Zhang. "Feature extraction of underwater targets using generalized S-transform." *J. Comput. Appl.* 32 (2012): 280-282.

[2] Liu, L. X., S. H. Jiao, and T. Chen. "Detection and recognition of underwater target based on feature matching." *Modern Electronics Technique* 34, no. 4 (2014): 73-76.

[3] Liu, L. I. K., and M. Dun. "Algorithm for recognition of underwater small target based on shape characteristic." *Ship Sci. Technol* 34, no. 1 (2012).

[4] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).

[5] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788. 2016.

[6] Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263-7271. 2017.

[7] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pp. 21-37. Springer International Publishing, 2016.

[8] Dai, Jifeng, Yi Li, Kaiming He, and Jian Sun. "R-fcn: Object detection via region-based fully convolutional networks." *Advances in neural information processing systems* 29 (2016).

[9] He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn." In *Proceedings of the IEEE international conference on computer vision*, pp. 2961-2969. 2017.

[10] Muthana, Mahmoud, and Ahmed R. Nasser. "Using Dynamic Pruning Technique for Efficient Depth Estimation for Autonomous Vehicles." *Mathematical Modelling of Engineering Problems* 9, no. 2 (2022).

[11] Srikrishnan, A., Arun Raaza, Abishek B. Ebenezer, V. Rajendran, M. Anand, and S. Gopalakrishnan. "A Fast and Effective Method for Intrusion Detection using Multi-Layered Deep Learning Networks." *International Journal of Advanced Computer Science and Applications* 13, no. 12 (2022).

[12] Alowaidi, Majed. "Modified Intrusion Detection Tree with Hybrid Deep Learning Framework based Cyber Security Intrusion Detection Model." *International Journal of Advanced Computer Science and Applications* 13, no. 10 (2022).

[13] Ummadisetti, Ganesh Naidu, R. Thiruvengatanadhan, Satyala Narayana, and P. Dhanalakshmi. "Character level vehicle license detection using multi layered feed forward back propagation neural network." *Bulletin of Electrical Engineering and Informatics* 12, no. 1 (2023): 293-302.

[14] Santoso, Albertus Joko, and Raymond Erz Saragih. "Automatic Face Mask Detection Based on MobileNet V2 and Densenet 121 Models." *ICIC Express Letters* 16, no. 4 (2022): 433-440.

[15] Li, Xiu, Min Shang, Jing Hao, and Zhixiong Yang. "Accelerating fish detection and recognition by sharing CNNs with objectness learning." In *OCEANS 2016-Shanghai*, pp. 1-5. IEEE, 2016.

[16] Chen, Long, Feixiang Zhou, Shengke Wang, Junyu Dong, Ning Li, Haiping Ma, Xin Wang, and Huiyu Zhou. "SWIPENET: Object detection in noisy underwater images." *arXiv preprint arXiv:2010.10006* (2020).

[17] Wei, Xiangyu, Long Yu, Shengwei Tian, Pengcheng Feng, and Xin Ning. "Underwater target detection with an attention mechanism and improved scale." *Multimedia Tools and Applications* 80, no. 25 (2021): 33747-33761.

[18] Zeng, Lingcai, Bing Sun, and Daqi Zhu. "Underwater target detection based on Faster R-CNN and adversarial occlusion network." *Engineering Applications of Artificial Intelligence* 100 (2021): 104190.

[19] Chen, Lingyu, Meicheng Zheng, Shunqiang Duan, Weilin Luo, and Ligang Yao. "Underwater target recognition based on improved YOLOv4 neural network." *Electronics* 10, no. 14 (2021): 1634.

[20] Cao, Xiang, Lu Ren, and Changyin Sun. "Dynamic target tracking control of autonomous underwater vehicle based on trajectory prediction." *IEEE Transactions on Cybernetics* 53, no. 3 (2022): 1968-1981.

[21] Huang, Hai, Hao Zhou, Xu Yang, Lu Zhang, Lu Qi, and Ai-Yun Zang. "Faster R-CNN for marine organisms detection and recognition using data augmentation." *Neurocomputing* 337 (2019): 372-384.

[22] Chen, Zhe, Hongmin Gao, Zhen Zhang, Helen Zhou, Xun Wang, and Yan Tian. "Underwater salient object detection by combining 2D and 3D visual features." *Neurocomputing* 391 (2020): 249-259.

[23] Lin, Wei-Hong, Jia-Xing Zhong, Shan Liu, Thomas Li, and Ge Li. "Roimix: Proposal-fusion among multiple images for underwater object detection." In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2588-2592. IEEE, 2020.

[24] Song, Pinhao, Pengteng Li, Linhui Dai, Tao Wang, and Zhan Chen. "Boosting R-CNN: Reweighting R-CNN samples by RPN's error for underwater object detection." *Neurocomputing* 530 (2023): 150-164.

[25] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

[26] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

[27] Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "Swin transformer: Hierarchical vision transformer using shifted windows." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012-10022. 2021.

[28] Hou, Qibin, Daquan Zhou, and Jiashi Feng. "Coordinate attention for efficient mobile network design." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13713-13722. 2021.

[29] Pan, Xuran, Chunjiang Ge, Rui Lu, Shiji Song, Guanfu Chen, Zeyi Huang, and Gao Huang. "On the integration of self-attention and convolution." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 815-825. 2022.

[30] Loshchilov, Ilya, and Frank Hutter. "Decoupled weight decay regularization." *arXiv preprint arXiv:1711.05101* (2017).

[31] Zhu, Xizhou, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. "Deformable detr: Deformable transformers for end-to-end object detection." *arXiv preprint arXiv:2010.04159* (2020).

[32] Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal loss for dense object detection." In *Proceedings of the IEEE international conference on computer vision*, pp. 2980-2988. 2017.

[33] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature pyramid networks for object detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117-2125. 2017.

[34] Qiao, Siyuan, Liang-Chieh Chen, and Alan Yuille. "Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10213-10224. 2021.

[35] Tian, Zhi, Chunhua Shen, Hao Chen, and Tong He. "Fcos: Fully convolutional one-stage object detection." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9627-9636. 2019.

[36] Zhou, Xingyi, Dequan Wang, and Philipp Krähenbühl. "Objects as points." *arXiv preprint arXiv:1904.07850* (2019).

A Computational Prediction Model of Blood-Brain Barrier Penetration Based on Machine Learning Approaches

Deep Himmatbhai Ajabani
Application Developer, Lead, Source InfoTech Inc., Atlanta, Georgia, United States

***Abstract*—Within the field of medical sciences, addressing brain illnesses such as Alzheimer's disease, Parkinson's disease, and brain tumors poses significant difficulties. Despite thorough investigation, the search for truly successful neurotherapies continues to be challenging to achieve. The blood-brain barrier (BBB), which is currently a major area of research, restricts the passage of medicinal substances into the central nervous system (CNS). It is crucial in the field of neuroscience to create drugs that can effectively cross the blood-brain barrier (BBB) and treat cognitive disorders. The objective of this study is to improve the accuracy of machine learning models in predicting BBB permeability, which is a critical factor in medication development. In recent times, a range of machine learning models such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Artificial Neural Networks (ANN), and Random Forests (RF) have been utilized for BBB. By employing descriptors of varying dimensions (1D, 2D, or 3D), these models demonstrate the potential to make precise predictions. However, the majority of these studies are biased to the nature of datasets. To accomplish our objective, we utilized three BBB datasets for training and testing our model. The Random Forest (RF) model has shown exceptional performance when used on larger datasets and extensive feature sets. The RF model attained an overall accuracy of 90.36% with 10-fold cross-validation. Additionally, it earned an AUC of 0.96, a sensitivity of 77.73%, and a specificity of 94.74%. The assessment of an external dataset resulted in an accuracy rate of 91.89%, an AUC value of 0.94, a sensitivity rate of 91.43%, and a specificity rate of 92.31%.**

Keywords*—Central Nervous System (CNS); Blood-Brain Barrier (BBB); Machine Learning (ML); Simplified Molecular Input Line Entry System (SMILES); Support Vector Machine (SVM); K-Nearest Neighbor (KNN); Logistic Regression (LR); Multi-Layer Perceptron (MLP); Light Gradient Boosting Machine (LightGBM); Random Forest (RF)

I. Introduction

In the last few decades, human brain diseases like tumors in the brain, dementia, Alzheimer, and other brain disorders are the most common fastest-growing issue nowadays that causes disability in humans and received the most attention from the research community in medical sciences. As there are no effective treatments have been made by neurotherapist to treat these kinds of serious diseases. Almost all macro and small molecule drugs are blocked by a barrier named the BBB [1]. The BBB is the most important key point in the treatment of brain diseases as it forcibly prevents the drugs from crossing this barrier and enters the CNS [2]. Many ML and Deep learning techniques have been made in the past and till now most researchers have been working on this problem of BBB permeability but still, the question arises on the performance of models and their precise results for drug formation in pharmaceutics [3], [4].

As the name indicates BBB, it is the barrier between blood and the brain. The barrier is made of endothelium cells [5] which can prevent large and even small molecules from entering the CNS [6]. It only allows some specific molecules like water molecules and some lipid-soluble to cross the barrier [7]. The BBB is divided into two classes labeled BBB+ and BBB-. The BBB+ shows the higher permeability and the BBB- shows the lower permeability respectively [8]. Developing a classification model requires a piece of detailed information and a complete understanding of issues or problems regarding BBB permeability. These issues are mainly caused by the selection of algorithms and the dataset on which these computations are performed. The problems faced in algorithms include their lower coverage, overfitting w.r.t dataset, and lower accuracy scores while predicting the molecular compounds i.e., (BBB-). The problems raised regarding datasets are duplication of compounds or improper class label distribution in the BBB dataset, which is a serious cause of inaccurate results [9].

In BBB there are molecular descriptors used as features in the dataset. The definition of molecular descriptor states that the transformation of chemical compounds by applying mathematical procedure converts these compounds into standardized numeric information that can be used for further experiments [10]. Molecular descriptors encompass various characteristics of molecules, such as their weight, amount of carbon atoms, and hydrogen bonds. The literature review primarily focused on the discussion of various classes of molecular descriptors, namely one dimension, two dimensions, and three dimensions. The representation of molecular descriptors is categorized into several kinds. Numerous classes have been discussed in the existing body of scholarly literature. The classes were categorized into three distinct groups, namely 1D, 2D, and 3D.

One-dimensional molecular descriptors, such as the number of certain atoms and molecular weights, are utilized to express the attributes of molecules [11]. The presentation of structural information is accomplished by the utilization of 2D molecular descriptors. It is computed from the 2D molecular structure like the number of donors in the H bond, the number of C6H6 rings, etc. [11]. The structural information is represented by 3D molecular descriptors like a positive partial charge structure of solvent-accessible surface area [11].

In this study, the main focus is given to the diversity of datasets in terms of size and nature of datasets, further applying a variety of machine learning algorithms. The novelty of the work is the generation of chemical features from SMILES and testing them as unseen data for the best model. Further, we have tested machine learning algorithms with different hyperparameters and chosen the best hyperparameter for each algorithm that was missing in the previous literature. In previous studies, experiments were conducted with default hyperparameters [7], [12-15]. Also, the model is evaluated on several different evaluation metrics to validate the performance of the best-chosen model.

II. Related Work

BBB is an up-and-coming research area that is widely used in the formation of drug discovery. In the last decades, several approaches to the BBB have been proposed. These approaches are based on ML algorithms and have followed their method of technique. Permitting the literature study, there were several approaches have been proposed which have their methods and techniques. These techniques vary with the number of compounds used and the selection of important features related to these chemical compounds. So, the proposed system will give outcomes in terms of results of model accuracy, sensitivity, specificity, and the robustness of testing scores.

Dai et al. 2021 [1] proposed a feature representation in sequential-based prediction for BBB peptides. In this study, 16 classes of peptide sequence feature descriptors have been used. For finding the best solution three-step method was used. In the three-step model, features were selected based on the F1 score, and Spearman's rank-order correlation and a sequential forward selection strategy were implemented. In this study, many ML models were compared i.e., ERT, XGB, LR, MLP, RF, and SVM. While comparing the results of each model the LR has the best prediction ability to gain an overall AUC and AUPR score of 0.87 with 10-fold cross-validation. But dataset contains only 119 BBPs compound datasets having only seven features for classification and mainly focusing on peptide-based molecular compounds. On the contrary, Zou 2022 [2] uses the physicochemical properties of amino acids and through these amino properties, the author identifies blood-brain barrier peptides (BBPs) and also applies the features fusion method. In this research, SVM was implemented on a dataset that represents peptide sequences based on 100 samples from BBPs, and 100 samples from non-BBPs were used, together with 10 physicochemical characteristics descriptors. For the selection of discriminative features, the Fisher algorithm was used. The highest accuracy, specificity, and sensitivity achieved by the model on the training dataset is 100%, while MCC and AUC are also 1.00, while on the independent dataset, it was 89.47%. The limitation of this work is limited samples were used and they just employed the correlation information between two different types of physicochemical properties. Also, there is a lack of biological experiments to validate the predicted results.

Similarly, Shaker et al. 2021 [7] proposed a LightGBM algorithm that was implemented on a 7162 compounds dataset with BBB permeability in which 5453 BBB+ and 1709 BBB- class with 1119 molecular descriptors of SMILES format. 10-fold cross-validation was implemented after splitting the dataset into 10% testing and 90% training and the results show an accuracy score of 0.89, specificity of 0.77, sensitivity of 0.93, and area under the curve of 0.94 respectively. However, the accuracy can be improved in the future by testing other ML models as it is critical to decide which molecular compound can penetrate from CNS through BBB. However, the use of these many features increases the complexity of the models. Therefore, Alsenan et al. 2021 [9] proposed a model that used the Kernel Principal Component Analysis (KPCA) method for finding descriptors. The author also compares the deep learning (DL) models with ML models and comes with the result that deep learning models show more accurate results than ML models. The FFDNN and CNN achieve accuracy of 100%, specificity of 98.11 and 99.87, and sensitivity of 96.78 and 98.76 respectively. The AUC was also calculated which was 97.7 and 99.71 and Matthew's correlation coefficient was 95.55 and 92.85 respectively. However, the dataset was composed of 2500 molecular compounds with 6,394 molecular descriptors which are small datasets as a large dataset has a direct impact on accuracy and only focuses on the KPCA feature extraction technique.

Furthermore, various variants of ML models are tested by Kumar et al. 2021 [12]. The author proposed an RF-based method for the prediction of the BBB by using chemical peptides. Different algorithms were implemented i.e., DT, RF, LR, KNN, GNB, XGB, and SVC. Three datasets were used in this study i.e., dataset-1 had 269 B3PPs and CPPs respectively. Dataset-2 was having 269 B3PPs and non-B3PPs respectively, while dataset-3 was having 269 B3PPs and 2690 non-B3PPs. The highest accuracy, specificity, and sensitivity were achieved by RF, which is 85.08, 85.08, and 86.97 respectively. Matthew correlation and AUC are also calculated which are 0.51 and 0.93. But the author collected only 465 peptides from the B3Pdb database which is a small dataset of cell-penetrating peptides with 80 selected features.

Also, Liu et al. 2021 [13] proposed the SMOTE technique on 1757 chemical compounds, and the feature descriptors were produced by PaDEL-Descriptor software for nine molecular fingerprints and 2D and 3D descriptors on five-fold cross-validation with 100 iterations. Three algorithms were implemented i.e., SVM, RF, and XGBoost from which RF shows the higher scores in terms of accuracy of 0.910, specificity of 0.867, sensitivity of 0.927, and AUC of 0.957 respectively. But there are a smaller number of descriptors used and this model is not a quantitative approach to identifying which BBB chemical compound can penetrate or not. In comparison, Shi et al. 2021 [14] in this approach, 2354 drug molecules of SMILES format were used with 33 molecular features. 10-fold cross-validation was used and six types of methods were used for training of imbalance dataset i.e., Upsampling, RUS, Weight parameter, SMOTE, SMOTECENN, and ADASYN. The results clearly show that XGBoost outperforms other approaches in terms of precision 0.92, recall 0.96, F1-score 0.94, Accuracy 0.95, specificity

0.93, sensitivity 0.98, and AUC 0.98 respectively. It is worth mentioning that, using too many resampling methods can lead to overfitting and inaccuracy. So, it may harm the model's outcomes.

Saber et al. 2020 [15] proposed a comparative approach to ML algorithms. The algorithms that are implemented in this research study are SVM with linear, polynomial, radial basis function kernels, LDA and QDA, and KNN. The author concludes that a genetic algorithm with SVM outperforms other approaches. It shows an accuracy of 96.23, a specificity of 86.67, and a sensitivity of 98.45. All algorithms compiled on 1593 drug compounds and eight molecular descriptors were generated by sequential feature selection and genetic algorithm. There is a lack of a greater number of features and training the model on fewer features has a great impact on the outcomes. Similar dataset dimension biases also happened in the study proposed by Ciura et al. 2020 [16]. The authors suggested a technique that focuses on micellar electrokinetic chromatography and has 50 2D and 3D molecular descriptors from a collection of market available 45 chemical drugs. MLR and SVM were implemented on a given dataset and showed the same results for prediction, by showing the same results of RMSE and cross-validation of 0.310 and 0.314 respectively. But if a large dataset and a greater number of features were applied this will affect the results as model accuracy directly depends on the size of the dataset.

Moreover, a study proposed by Singh et al. 2020 [17] comprised a novel validation approach of QSAR. In this approach, RF has been implemented on a 605 compounds dataset with 1444 molecular descriptors of 1D and 2D generated by PaDEL software 2.21. In the proposed methodology 10-fold cross-validation QSAR approach was used. Two types of thresholds were employed to divide the dataset. Specifically, threshold-1 was defined as (B/P>=0.6 classified as BBB+ and B/P<0.6 classified as BBB-), while threshold-2 was defined as (B/P>0.6 classified as BBB+ and B/P<0.3 classified as BBB-). Threshold-1 and threshold-2 attained precision of 86% and 87% accordingly. However, this study defined a specific range of thresholds to specify the classes of BBB and only focused on the QSAR approach may other techniques improved the results of the proposed model.

A few other researchers such as Radchenko et al. 2020 [18] implemented an artificial neural network on 529 molecular compounds datasets based on their LogBB values and 100 to 1000 descriptors were generated by using substructures of molecular compounds. The silico LogBB-based model used fragmental substructural descriptors representing the occurrence number of the various substructures. The results show that Q2 has a value of 0.815 and an RMSE of 0.318. However, this research work only concentrates on LogBB values of compounds with a small dataset of compounds. Saxena et al. 2019 [19] presented an ML model for permeability prediction of the BBB. In this study, SVM, KNN, RF, and NB were implemented in 1978 molecular compounds. Physicochemical characteristics, MACCS fingerprints, and substructure fingerprints were included in 1917 feature vectors. With an accuracy score of 96.77 percent, SVM with RBF kernel performs better as compared to other proposed ML techniques. However, the dataset used in this research study has a smaller number of chemical compounds which can affect the results. Roy et al. 2019 [20] proposed an approach SVM, KNN, gradient boost machine, and the statistical importance analysis method used to select 37 descriptors, and a generalized linear model was implemented on it. The results show that SVM surpasses other approaches with an accuracy of 96%, a sensitivity of 99%, a specificity of 87%, a precision of 96%, and an F1 score of 97% respectively. The dataset contains 1800 molecules and was divided into 75% training data and 25% testing data. Rui Miao et al. 2019 [21] proposed three clinical phenotypes data of 1000 molecular compounds were used. DL method, SVM with sigmoid, polynomial, radial basis kernel functions, KNN, and DT were implemented. The dataset was utilized for both training and testing with five-fold cross-validation. As 70 percent is used for training and 30 percent is used for testing. The author concludes that the deep learning method outperforms other ML algorithms in terms of area under curve, accuracy, and F1-score i.e., 98%, 97%, and 92% respectively. Saber et al. 2019 [22] implement SVM, ANN, and KNN models with 1593 drug compounds. For the selection of molecular features, a genetic algorithm was used which generated 8 descriptors. The highest overall accuracy was obtained with both Quadratic Discriminant Analysis and SVM classifiers at 96.23%. But this research study only focuses on ADMET characteristics of compounds, and it is not clear how well the system detects permeable compounds because of the small dataset.

By Analyzing the related work, it is observed that the above research have some drawbacks. First, most of the researchers target datasets having a smaller number of molecular compounds. Second, the number of features used in the research is too small, and vice versa.

III. Methods and Methodology

A. Algorithm for Proposed Study

The algorithm for the proposed study was given below which inputs the dataset and applies preprocessing techniques to the given dataset. After preprocessing the dataset is divided into 90% training and 10% testing purposes and for each molecular compound feature values were generated. By applying the ML models on preprocessed datasets if the molecular compound belongs to class BBB+ it would be updated to the permeable list and else the molecular compound belongs to class BBB- it would be updated to the non-permeable list. For validation of our model, we test it on a test dataset and evaluate these results by using an evaluation matrix.

Algorithm 1: Algorithm for Prediction of BBB Permeability

```
Input: BBB Dataset
Output: List of compounds into permeable and non-permeable
1  Initialization
2  Input dataset
3  Refining an initiated dataset
4  Selection of 90% dataset for training data
5  Selection of 10% dataset for testing data
6  Filtration of data for required features
```

```
Applying ML models
For (i=1; i<= n; i++),
    If (Comp(i) are permeable == Yes)
        Updating permeable list
    Else
        Updating non-permeable list
    End If
```

B. Results of Flow Charts

The flow chart of the proposed methodology is discussed in Fig. 1. The dataset contains molecular compounds loaded for the filtration process. After filtration, the dataset was divided into 90% training and 10% testing. After the division of the dataset feature extraction process was applied in which 1D and 2D features were extracted for each compound. The most well-known ML models i.e., SVM, KNN, LR, ANN, and RF were applied to each chemical compound. ML models classify the dataset into two class labels i.e., BBB- (0) non-penetrating list and BBB+ (1) penetrating list. For evaluation and validation of results accuracy, specificity, sensitivity, precision, and recall, the F1-score is applied.

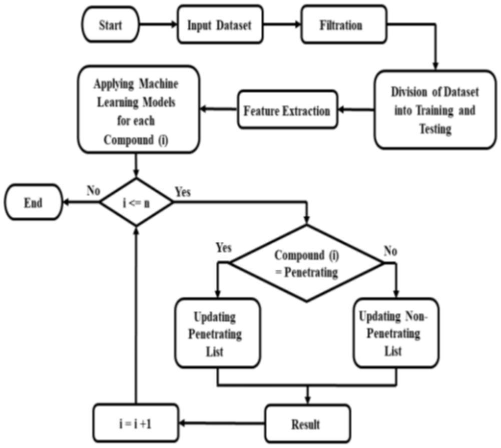

Start
Input Dataset
Filtration
Division of Dataset into Training and Testing
Feature Extraction
Applying Machine Learning Models for each Compound (i)
No
i <= n
Yes
End
Yes
Compound (i) = Penetrating
No
Updating Penetrating List
Updating Non-Penetrating List
Result
i = i +1

Fig. 1. Flow chart of the BBB permeability prediction model.

C. Data Collection

The datasets used in the proposed study were collected from the online repository of the LightBBB [23] web server which was in SMILES format [24]. In these datasets, the compounds were grouped as BBB+ which belongs to class 1, and BBB- which belongs to class 0. There are numerous descriptors available for expressing the BBB permeability chemicals. It's crucial to pick efficient descriptors for model training to prevent overfitting and poor performance. The training dataset included chemical compounds with 1D and 2D descriptors for each compound after dataset pre-processing.

D. Data Preprocessing

Preprocessing is an essential task for each ML model. To obtain effective results preprocessing is to be done on the dataset. The accuracy of the model is directly impacted by the size of the dataset. In this research study, molecular compounds were compiled with the experimental BBB permeability which leads to compounds belonging to the class of BBB+ and BBB-. A SMILES format was used to prepare each molecule. The BBB+ belonged to class 1 and BBB- belongs to class 0. The dataset was preprocessed to remove duplicates inconsistent compounds and missing structural information data were also removed. The dataset was split into 90% training and 10% testing using ten-fold cross-validation, with each iteration of validation being repeated ten times. For testing purposes, the external dataset contains molecular compounds of which some compounds belong to BBB+ and BBB- classes.

E. Feature Set

The physical and chemical characteristics of substances were described using molecular descriptors as features. As a result, these aspects provide more information to create a reliable BBB model [7].

F. Machine Learning Models

In recent decades, there has been a numerous growth in ML models and most of all have been used in the prediction of the BBB. Some BBB prediction models show good performance with a high accuracy score. Therefore, it was a challenging task to develop an ML model for the prediction of BBB permeability as the dataset of BBB available in biological science gives the impression of being limited [19]. ML approaches are classified as supervised, unsupervised, or reinforced. There are many supervised learning techniques some of them are; SVM, KNN, LR, ANN, and RF are mainly used for classification or regression problems and some deep learning-based algorithms are used for the prediction of BBB. Scikit-learn, a Python-based toolkit, were used in the implementation of the model.

1) Support Vector Machine (SVM): The support vector machine (SVM) is a widely recognized approach in supervised learning, commonly employed for solving classification and regression problems. The proposal was initially forth during the decade of the 1990s and has since been effectively utilized within the fields of bioinformatics and computer-aided diagnosis. Therefore, the SVM classification model was utilized in this research work to analyze the BBB dataset. The support vector classifier (SVC) is implemented using the support vector machines toolkit. The algorithm typically accommodates the supplied attributes' points of information and identifies the optimal hyperplane for classifying the data into two distinct categories [12]. The SVM model's effectiveness over traditional methods can be attributed to its inherent structure and the risk management philosophy it employs. Multiple kernel functions, such as polynomial, linear, radial basis function, and sigmoid, are utilized in Support Vector Machines (SVM) to facilitate the transformation of data into a higher-dimensional space where a distinct separation between classes can be achieved [20]. This research paper examines the performance of the Support Vector Machine (SVM) algorithm, specifically utilizing a linear kernel function, in the context of classification problems.

2) K-Nearest Neighbors (KNN): The k-nearest neighbor (KNN) approach is an example of supervised machine learning that may be applied to both classification and regression tasks. The technique, which was introduced in 1951, has gained significant popularity over the years as a reliable method for predicting drug penetration and blood-brain barrier (BBB) permeability. Unlabeled datasets are classed through the assignment of a class based on the similarity to neighboring data points. The K-nearest neighbors (KNN) algorithm computes the distances among the information points, namely the feature values, using metrics such as Euclidean distance or Manhattan distance. In the context where a desirable value of k is sought, it is necessary to evaluate multiple neighboring values. The decision to choose one of these neighbors has a significant influence on the general efficacy of the prediction system that is being constructed. Typically, the value of k is limited to an integer not exceeding 20 [19].

3) Logistic Regression (LR): Logistic regression is one of the most popular statistical models that uses a logistic function for binary dependent variables [14]. This algorithm comes under the tree of supervised machine techniques. LR is like linear regression as linear regression is used for regression problems and LR is used for solving classification problems.

4) Random Forest (RF): Random Forest (RF) is a machine-learning technique that is based on decision trees. The bootstrap resampling approach allows for the extraction of multiple samples from the original set of data. Following the choice of specimens, a prediction decision-making structure was constructed for each sample, which was subsequently aggregated through a voting mechanism to obtain the ultimate outcomes. Random Forest (RF) can be utilized to address both classification and regression problems. The primary advantage of the RF model is its ability to mitigate errors resulting from asymmetrical data during training, particularly when there is a substantial disparity between the two types of class compounds. Additionally, this model has superior performance in mitigating the overfitting phenomena, as well as exhibiting enhanced capability in effectively addressing outliers and noisy data [13].

5) Multi-Layer Perceptron (MLP): A multi-layer perceptron (MLP) refers to an artificial neural network characterized by a forward architecture, wherein it transforms a given set of input vectors into a corresponding set of output vectors. The MLP can be conceptualized as a directed graph including multiple tiers of nodes. The subsequent layer is interconnected with the preceding layer. Every node, except the input node, represents a neuron that possesses a non-linear activation function. [16].

6) Light Gradient Boosting Machine (LGBM): Gradient boosting decision trees are a popular machine learning algorithm that combines the strengths of decision trees and gradient boosting. This algorithm iteratively builds an ensemble of weak decision trees, where There are various manifestations of trees, one of which is LightGBM (GBDT). This technique is commonly employed for classification, regression, and efficient parallel training. The LightGBM algorithm is widely recognized as a rapid and efficient variation of the Gradient Boosting Decision Tree (GBDT) technique. The proposed approach involves partitioning the tree into individual leaves and thereafter identifying the leaf that exhibits the highest delta loss. Hence, under the LightGBM framework, the leaf-wise approach can minimize loss to a greater extent compared to the level-wise strategy when expanding on the identical leaf. [7]. The description of all the parameters applied to ML models is discussed below in Table I.

G. *Description of Datasets*

ML models are implemented on three BBB datasets. These datasets have two classes i.e., class 0 belongs to (BBB-) which specifies non-permeable compounds to BBB and class 1 belongs to (BBB+) permeable compounds to BBB. All the molecular compounds were in SMILES format. The datasets were randomly divided into 90% training and 10% testing data on which ML models were trained.

Dataset 1 contains 1072 (317 BBB+ and 755 BBB-) molecular compounds in SMILES format with a variety of 196 1D descriptors generated from RDKit library [25] which is a Python built-in library mainly used for molecular descriptors. The test dataset contains 266 molecular compounds with 196 1D descriptors that are extracted for each compound.

TABLE I. DESCRIPTION OF PARAMETERS FOR MACHINE LEARNING MODELS

Name	Value	Description
Kernel	Linear	Defines the type of kernel to be used in the algorithm
Random_state	None	Controls the creation of pseudo-random numbers used to shuffle the data used to calculate probabilities.
N_neigbors	7	The numbers of neighbours that k-neighbors queries will by default utilize.
Metric	Minkowski	Metric to employ for distance calculations that, when p = 2, yields the usual Euclidean distance.
P	2	Minkowski metric's power parameter
Solver	liblinear	The optimization problem's algorithm.
Random_state	None	Used when Solver = liblinear
Hidden_layer_sizes	(8, 8, 8)	The number of neurons in the ith hidden layer is represented by the ith element.
Activation	Relu	The buried layer rectified linear unit function's activation function gives the result f(x) = max (0, x)
Solver	Adam	A stochastic gradient-based optimizer is referred to in the solution for weight optimization.
Max_iter	2000	The maximum number of iterations.
n_estimator	100	It means how many numbers of trees can be generated.
Max_depth	None	The depth of trees.
Min_sample_split	2	The needed minimum number of samples
Random_state	42	It means how many times the function calls for the same instance.
Max_features	Sqrt	It means max_features= sqrt(n_features)

Dataset 2 contains 7162 (5453 BBB+ and 1709 BBB-) molecular compounds in SMILES format with 1119 1D and 2D descriptors that are extracted for each compound. The test dataset contains 74 (39 BBB+ and 35 BBB-) molecular compounds with 1119 1D and 2D descriptors that are extracted for each compound [23]. These features were generated using Dragon software (version 7.0.10) [26].

Dataset 3 was constructed by adding more chemical compounds in dataset 2 which contains 9230 (6852 BBB+ and 2378 BBB-) molecular compounds in SMILES format with 1119 1D and 2D descriptors that are extracted for each compound. The test dataset contains 74 (39 BBB+ and 35 BBB-) molecular compounds with 1119 1D and 2D descriptors that are extracted for each compound. These features were generated using Dragon software (version 7.0.10).

H. Evaluation Matrices

1) Confusion matrix: The utilization of the confusion matrix is frequently observed in machine learning to assess and illustrate the performance of algorithms in supervised classification tasks. The matrix in question is a square matrix whereby the rows correspond to the actual class and the columns correspond to the predicted class. The confusion matrix establishes a quantitative assessment of the concordance between observed and forecasted data.

2) Sensitivity: The sensitivity is defined as the percentage of chemical compounds that the model properly classifies as BBB+ [9] and it is calculated by the given formula as shown in Eq. (1).

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (1)$$

3) Specificity: The specificity is defined as the percentage of chemical compounds that the model properly classifies as BBB- [9] and it is calculated by the given formula as shown in Eq. (2).

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (2)$$

4) Accuracy: The accuracy shows the overall performance of the model [9] and it is calculated by the given formula as shown in Eq. (3).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (3)$$

5) Receiving Operating Characteristics (ROC): The model was graphically evaluated using an ROC curve [27], which is a highly efficient approach for determining how well the model can accurately distinguish between classes [7].

6) AUC: The AUC is used to assess how well the classifier separates the classes by calculating the area under the ROC curve and its output will always be between 0 and 1 [9].

IV. Results and Discussion

Dataset 1 contains 1072 chemical compounds with 196 1D descriptors on which ML models were trained. The dataset was divided into 90% for training and 10% for testing with 10-fold cross-validation and 10 times iterated the whole process. The results are demonstrated below in Table II.

On cross-validation, the training dataset contains 964 chemical compounds whereas the testing dataset contains 108 chemical compounds. The results on Dataset 1 show that we achieved an overall accuracy of the RF of 93.52, an AUC of 0.97, a sensitivity of 95.95, and a specificity of 88.24 on 10-fold cross-validation. The higher AUC value indicates that our model has a high level of accuracy in predicting BBB permeability and is suitable for use in BBB prediction. In contrast with other ML models, the RF model outperforms other ML models as shown in Fig. 2. The results of ML models are demonstrated by using ROC Curve for cross-validation on dataset 1 as shown in Fig. 2.

For validation of the models, we test the ML models on an external dataset 1. Fig. 3 shows ROC curve of ML models for cross-validation of dataset 1. The RF model shows an accuracy of 78.38, an AUC of 0.83, a sensitivity of 94.29, and a specificity of 64.1. Comparing RF results with other ML models clearly shows that RF outperforms in the prediction of BBB permeability compounds.

The LightBBB dataset contains 7162 molecular compounds and was divided into 90% training and 10% testing. After the division of the dataset feature extraction process was applied. The LightBBB dataset contains 1119 1D and 2D descriptors extracted for each chemical compound. These descriptors were generated using Dragon software (version 7.0.10). The most well-known ML models i.e., the SVM, KNN, LR, ANN, and RF were applied to each chemical compound. ML models classify the dataset into two class labels i.e., BBB- (0) non-penetrating list and BBB+ (1) penetrating list. For evaluation and validation of results accuracy, specificity, sensitivity, precision, and recall, the F1-score has been computed. The results are demonstrated below in Table III.

TABLE II. Cross-Validation Results of ML Models on Dataset 1

Models	AUC	Specificity	Sensitivity	Accuracy
SVM	0.91	88.24	83.78	85.19
KNN	0.95	94.12	81.08	85.19
LR	0.91	85.29	85.14	85.19
MLP	0.91	76.47	79.73	78.07
LGBM	0.97	88.24	94.59	92.59
RF	0.97	88.24	95.95	93.52

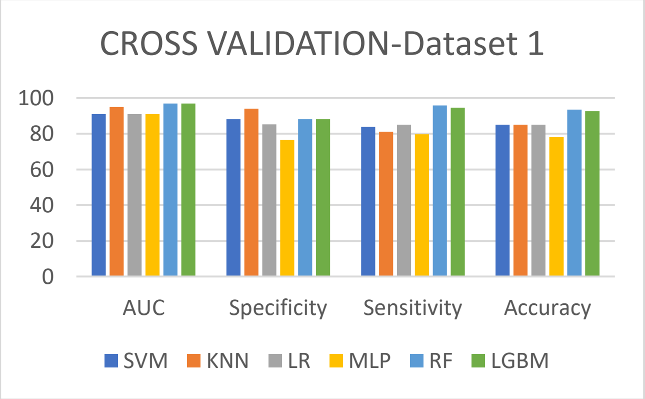


Fig. 2. Performance of ML models on cross validation dataset 1.

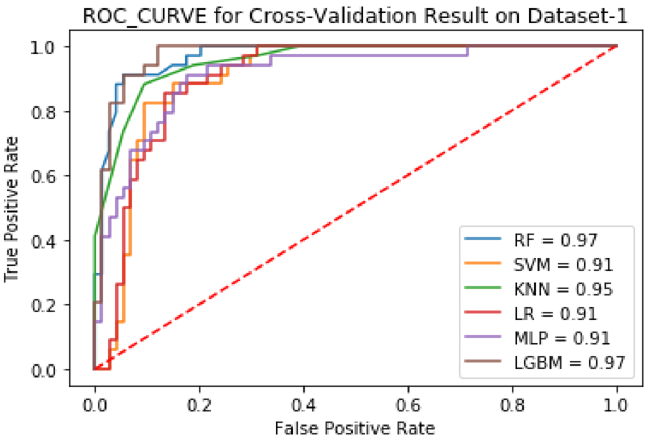


Fig. 3. ROC curves of ML models for cross validation on dataset 1.

TABLE III. CROSS-VALIDATION RESULTS OF ML MODELS ON DATASET 2

Models	AUC	Specificity	Sensitivity	Accuracy
SVM	0.90	93.49	73.78	88.98
KNN	0.92	96.02	62.8	88.42
LR	0.91	94.03	73.78	89.4
MLP	0.92	93.41	75.44	89.12
LGBM	0.94	0.77	0.93	89
RF	0.95	95.12	75.0	90.52

On cross-validation, the training dataset contains 6445 chemical compounds whereas the testing dataset contains 717 chemical compounds. The results on Dataset 2 show that we achieved an overall accuracy of the RF of 90.52, an AUC of 0.95, a sensitivity of 75.0, and a specificity of 95.12 on 10-fold cross-validation. The higher AUC value indicates that the RF model has a high level of accuracy in predicting BBB permeability and is suitable for use in BBB prediction. In contrast with other ML models, the RF model outperforms other ML models as shown in Fig. 4. The results of ML models are demonstrated by using the ROC Curve on cross-validation dataset 2 as shown in Fig. 5. For validation of the models, we test the ML models on an external dataset 2. The RF model shows an accuracy of 93.24, an AUC of 0.96, a sensitivity of 91.43, and a specificity of 94.87. Comparing RF results with other ML models clearly shows that RF outperforms in the prediction of BBB permeability compounds.

Dataset 3 contains 9230 molecular compounds and was divided into 90% training and 10% testing. After the division of the dataset feature extraction process was applied. The dataset contains 1119 1D and 2D features extracted for each chemical compound. The most well-known ML models i.e., the SVM, KNN, LR, ANN, and RF were applied to each chemical compound. ML models classify the dataset into two class labels. The results are demonstrated below in Table IV.

On cross-validation, the training dataset contains 8307 chemical compounds whereas the testing dataset contains 923 chemical compounds. The results on Dataset 3 show that we achieved an overall accuracy of the RF of 90.36, an AUC of 0.96, a sensitivity of 77.73, and a specificity of 94.74 on 10-fold cross-validation. In contrast with other ML models, the RF model outperforms other ML models as shown in Fig. 6. The results of ML models are demonstrated by using the ROC Curve on cross-validation dataset 3 as shown in Fig. 7. For validation of the models, we test the ML models on an external dataset 3. RF shows an accuracy of 91.89, an AUC of 0.94, a sensitivity of 91.43, and a specificity of 92.31. Comparing RF results with other ML models clearly shows that RF outperforms in the prediction of BBB permeability compounds.

While comparing our results with previously published BBB permeability prediction models it seems that our technique outperforms the existing methods. The uniqueness of our technique is the use of optimal hyperparameters and a high density of data. We compared the models by considering all the evaluation parameters i.e., AUC, specificity, sensitivity, and accuracy as shown in Table V.

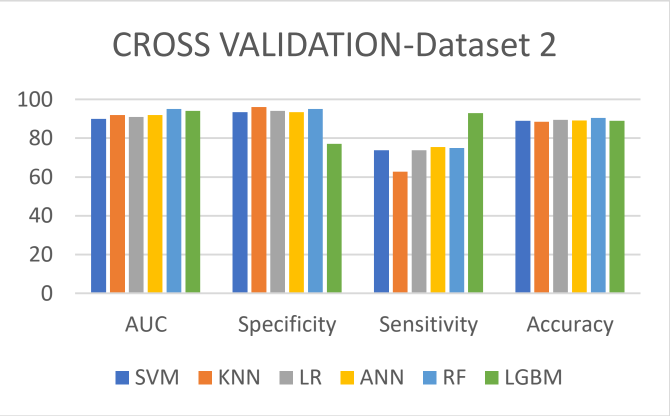


Fig. 4. Performance of ML models on dataset 2.

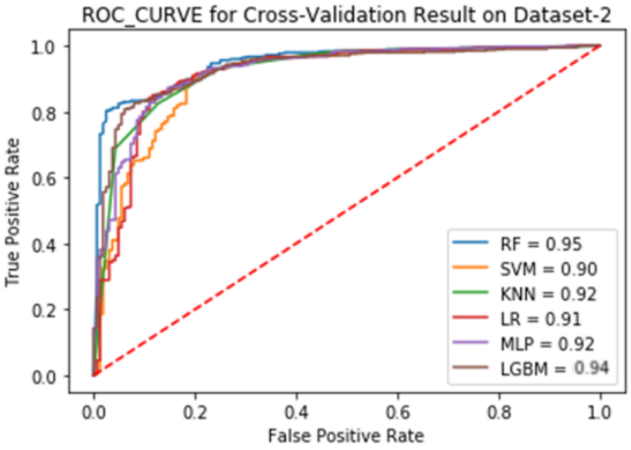


Fig. 5. ROC curves of ML models for cross validation on dataset 2.

TABLE IV. CROSS-VALIDATION RESULTS OF ML MODELS ON DATASET 3

Models	AUC	Specificity	Sensitivity	Accuracy
SVM	0.90	94.31	70.17	88.08
KNN	0.94	93.14	65.55	86.02
LR	0.92	93.72	68.49	87.22
MLP	0.96	91.94	85.09	90.25
LGBM	96	95.18	74.37	89.82
RF	0.96	94.74	77.73	90.36

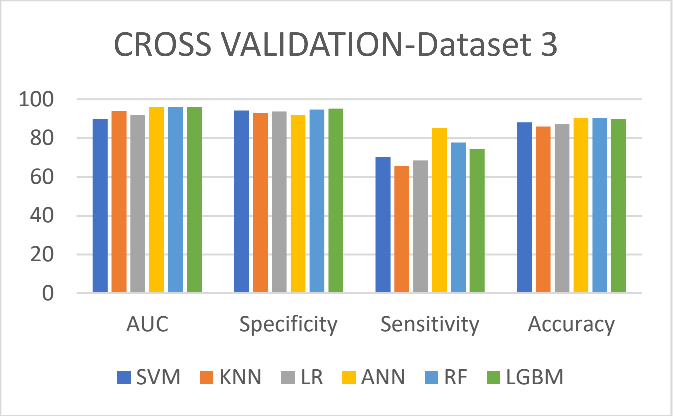


Fig. 6. Performance of ML models on dataset 3.

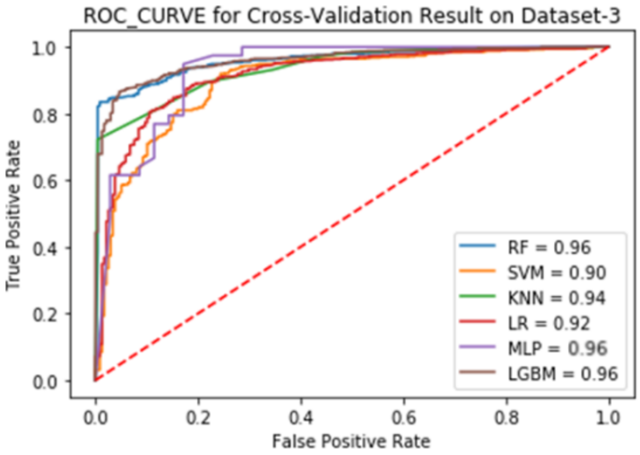


Fig. 7. ROC curves of ML models for cross validation on dataset 3.

TABLE V. COMPARISON OF ML MODELS WITH PREVIOUSLY PUBLISHED BBB MODELS

Reference	AUC	Specificity	Sensitivity	Accuracy
[7]	0.94	0.77	0.93	89%
[12]	0.93	0.85	0.86	85.08%
[13]	0.95	0.86	0.92	91%
[17]	-	0.71	0.92	87%
[21]	0.98	-	-	97%
[28]	0.90	0.83	0.98	94%
[29]	-	0.88	0.85	86%
[30]	0.78	-	-	82%
[31]	-	0.65	0.90	74.7%
[32]	-	0.80	0.82	81.5%
[33]	-	0.72	0.82	95%
[34]	-	0.80	0.72	83%
[35]	-	0.37	0.91	82.5%
[36]	-	0.79	0.84	82%
[37]	0.85	-	-	85%
Proposed Technique	**0.96**	**94.74**	**77.73**	**90.36%**

V. CONCLUSION

In the proposed study five machine learning models were applied to highly accurate small and large datasets with a larger number of features. The dataset is balanced and free from inconsistent and redundant data with accurate class labeling. On the contrary, other ML models were trained on a smaller dataset and fewer features, leading to differing accuracy levels but being unable to compensate for the variety of molecular components. The model uses 10-fold cross-validation with 10 iterations to assure correctness. The dataset contains molecular compounds and features. It was concluded that our ML model RF for the prediction of BBB penetration shows more accurate results on both small and large datasets than other ML algorithms.

The higher accuracy achieved by RF on dataset 1 is 93.52, with an AUC of 0.97, a sensitivity of 95.95, and a specificity of 88.24 on 10-fold cross-validation.

The higher accuracy achieved by RF on dataset 2 is 90.52, with an AUC of 0.95, a sensitivity of 75.0, and a specificity of 95.12 on 10-fold cross-validation. On testing our model on an external dataset RF shows an accuracy of 93.24, an AUC of 0.96, a sensitivity of 91.43, and a specificity of 94.87.

The higher accuracy achieved by RF on dataset 3 is 90.36, with an AUC of 0.96, a sensitivity of 77.73, and a specificity of 94.74 on 10-fold cross-validation. On testing our model on an external dataset RF shows an accuracy of 91.89, an AUC of 0.94, a sensitivity of 91.43, and a specificity of 92.31. The greater the value of AUC, the higher the accuracy of the model will be. Our model outperforms previously reported models.

VI. FUTURE WORK

To encourage future study, it may focus on the features of BBB datasets which can also be increased. It may also focus on applying feature extraction techniques for finding the most important features that were highly influential to the prediction compounds and how these models can be applied to treatments of the brain. Moreover, it may also focus on applying DL models to large datasets and comparing their outcomes with other ML models. Future work may intend to combine two techniques i.e., swarm algorithms with RF to obtain more precise results for this problem.

REFERENCES

[1] R. Dai et al., “BBPpred: Sequence-Based Prediction of Blood-Brain Barrier Peptides with Feature Representation Learning and Logistic Regression,” J Chem Inf Model, vol. 61, no. 1, pp. 525–534, 2021, doi: 10.1021/acs.jcim.0c01115.Ren, Y., et al. (2019). "Data storage mechanism based on blockchain with privacy protection in wireless body area network." Sensors 19(10): 2395.

[2] H. Zou, “Identifying blood-brain barrier peptides by using amino acids physicochemical properties and features fusion method,” Peptide Science, vol. 114, no. 2, 2022, doi: 10.1002/pep2.24247.Ren, C., et al. (2020). "Achieving Near-Optimal Traffic Engineering Using a Distributed Algorithm in Hybrid SDN." IEEE Access 8: 29111-29124.

[3] W. M. Pardridge, “The blood-brain barrier: Bottleneck in brain drug development,” NeuroRx, vol. 2, no. 1, pp. 3–14, 2005, doi: 10.1602/NEURORX.2.1.3.

[4] A. G.-C. opinion in drug discovery & development and undefined 1999, “The design and molecular modeling of CNS drugs.,” europepmc.org, Accessed: Oct. 01, 2022. [Online]. Available: https://europepmc.org/article/med/19649956.

[5] N. Abbott, L. Rönnbäck, E. H.-N. reviews neuroscience, and undefined 2006, “Astrocyte–endothelial interactions at the blood–brain barrier,” nature.com, Accessed: Oct. 01, 2022. [Online]. Available: https://www.nature.com/articles/nrn1824.

[6] H. Davson, “History of the Blood-Brain Barrier Concept,” Implications of the Blood-Brain Barrier and Its Manipulation, pp. 27–52, 1989, doi: 10.1007/978-1-4613-0701-3_2.

[7] B. Shaker et al., “LightBBB: Computational prediction model of blood-brain-barrier penetration based on LightGBM,” Bioinformatics, vol. 37, no. 8, pp. 1135–1139, 2021, doi: 10.1093/bioinformatics/btaa918.

[8] B. Hendricks, A. Cohen-Gadol, J. M.-N. focus, and undefined 2015, “Novel delivery methods bypassing the blood-brain and blood-tumor barriers,” thejns.org, doi: 10.3171/2015.1.FOCUS14767.

[9] S. Alsenan, I. Al-Turaiki, and A. Hafez, “A Deep Learning Approach to Predict Blood-Brain Barrier Permeability,” PeerJ Computer Science, vol. 7. pp. 1–26, 2021. doi: 10.7717/peerj-cs.515.

[10] M. C. Hutter, “Molecular Descriptors for Chemoinformatics (2nd ed.). By Roberto Todeschini and Viviana Consonni.,” ChemMedChem, vol. 5, no. 2, pp. 306–307, Feb. 2010, doi: 10.1002/CMDC.200900399.

[11] H. Hong et al., “Mold2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics,” J Chem Inf Model, vol. 48, no. 7, pp. 1337–1344, 2008, doi: 10.1021/CI800038F.

[12] V. Kumar, S. Patiyal, A. Dhall, N. Sharma, and G. P. S. Raghava, “B3pred: A random-forest-based method for predicting and designing blood–brain barrier penetrating peptides,” Pharmaceutics, vol. 13, no. 8, 2021, doi: 10.3390/pharmaceutics13081237.

[13] L. Liu et al., “Prediction of the Blood-Brain Barrier (BBB) Permeability of Chemicals Based on Machine-Learning and Ensemble Methods,” Chem Res Toxicol, vol. 34, no. 6, pp. 1456–1467, 2021, doi: 10.1021/acs.chemrestox.0c00343.

[14] Z. Shi, Y. Chu, Y. Zhang, Y. Wang, and D. Q. Wei, “Prediction of blood-brain barrier permeability of compounds by fusing resampling strategies and extreme gradient boosting,” IEEE Access, vol. 9, pp. 9557–9566, 2021, doi: 10.1109/ACCESS.2020.3047852.

[15] R. Saber, R. Mhanna, and S. Rihana, “A machine learning model for the prediction of drug permeability across the Blood-Brain Barrier: a comparative approach,” pp. 1–15, 2020.

[16] K. Ciura, S. Ulenberg, H. Kapica, P. Kawczak, M. Belka, and T. Bączek, “Assessment of blood–brain barrier permeability using micellar electrokinetic chromatography and P_VSA-like descriptors,” Microchemical Journal, vol. 158, p. 105236, 2020, doi: 10.1016/j.microc.2020.105236.

[17] M. Singh, R. Divakaran, L. S. K. Konda, and R. Kristam, “A classification model for blood brain barrier penetration,” J Mol Graph Model, vol. 96, p. 107516, 2020, doi: 10.1016/j.jmgm.2019.107516.

[18] E. v. Radchenko, A. S. Dyabina, and V. A. Palyulin, “Towards Deep Neural Network Models for the Prediction of the Blood-Brain Barrier Permeability for Diverse Organic Compounds,” Molecules, vol. 25, no. 24, 2020, doi: 10.3390/molecules25245901.

[19] D. Saxena, A. Sharma, M. H. Siddiqui, and R. Kumar, “Blood Brain Barrier Permeability Prediction Using Machine Learning Techniques: An Update,” Curr Pharm Biotechnol, vol. 20, no. 14, pp. 1163–1171, Aug. 2019, doi: 10.2174/1389201020666190821145346.

[20] D. Roy, V. K. Hinge, and A. Kovalenko, “To Pass or Not to Pass: Predicting the Blood-Brain Barrier Permeability with the 3D-RISM-KH Molecular Solvation Theory,” ACS Omega, vol. 4, no. 16, pp. 16774–16780, 2019, doi: 10.1021/acsomega.9b01512.

[21] R. Miao, L. Y. Xia, H. H. Chen, H. H. Huang, and Y. Liang, “Improved Classification of Blood-Brain-Barrier Drugs Using Deep Learning,” Sci Rep, vol. 9, no. 1, pp. 1–11, 2019, doi: 10.1038/s41598-019-44773-4.

[22] R. Saber, S. Rihana, and R. Mhanna, “In silico and in vitro Blood-Brain Barrier models for early stage drug discovery,” International Conference on Advances in Biomedical Engineering, ICABME, vol. 2019-Octob, pp. 1–4, 2019, doi: 10.1109/ICABME47164.2019.8940222.

[23] “LightBBB.” http://bioanalysis.cau.ac.kr:7030/ (accessed Jul. 29, 2022).

[24] D. W.-J. of chemical information and computer and undefined 1988, “SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules,” ACS Publications, vol. 28, no. 1, pp. 31–36, Feb. 1988, doi: 10.1021/ci00057a005.

[25] “RDKit.” Accessed: Oct. 01, 2022. [Online]. Available http://www.rdkit.org/.

[26] A. Mauri, A. Srl, V. Consonni, M. Pavan, and R. Todeschini, “Dragon software: An easy approach to molecular descriptor calculations,” researchgate.net, Accessed: Oct. 01, 2022. [Online]. Available: https://www.researchgate.net/profile/Andrea-Mauri-5/publication/216208341_DRAGON_software_An_easy_approach_to_molecular_descriptor_calculations/links/5da5c4b692851caa1ba601d4/DRAGON-software-An-easy-approach-to-molecular-descriptor-calculations.pdf.

[27] A. B.-P. recognition and undefined 1997, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” Elsevier, Accessed: Oct. 01, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320396001422.

[28] Z. Wang et al., “In Silico Prediction of Blood–Brain Barrier Permeability of Compounds by Machine Learning and Resampling Methods,” ChemMedChem, vol. 13, no. 20, pp. 2189–2201, 2018, doi: 10.1002/cmdc.201800533.

[29] “A simple method to predict blood-brain barrier permeability of drug-like compounds using classification trees,” ingentaconnect.com, Accessed: Oct. 01, 2022. [Online]. Available: https://www.ingentaconnect.com/content/ben/mc/2017/00000013/00000007/art00010.

[30] F. Plisson, A. P.-M. drugs, and undefined 2019, “Predicting blood–brain barrier permeability of marine-derived kinase inhibitors using ensemble classifiers reveals potential hits for neurodegenerative disorders,” mdpi.com, Accessed: Oct. 01, 2022. [Online]. Available: https://www.mdpi.com/403028.

[31] P. Crivori, G. Cruciani, … P. C.-J. of medicinal, and undefined 2000, “Predicting blood− brain barrier permeation from three-dimensional molecular structure,” ACS Publications, Accessed: Oct. 01, 2022. [Online]. Available: https://pubs.acs.org/doi/abs/10.1021/jm990968+.

[32] S. Doniger, T. Hofmann, and J. Yeh, “Predicting CNS permeability of drug molecules: Comparison of neural network and support vector machine algorithms,” Journal of Computational Biology, vol. 9, no. 6, pp. 849–864, 2002, doi: 10.1089/10665270260518317.

[33] I. F. Martins, A. L. Teixeira, L. Pinheiro, and A. O. Falcao, “A Bayesian approach to in Silico blood-brain barrier penetration modeling,” J Chem Inf Model, vol. 52, no. 6, pp. 1686–1697, Jun. 2012, doi: 10.1021/CI300124C.

[34] A. Guerra, J. A. Páez, and N. E. Campillo, “Artificial neural networks in ADMET modeling: Prediction of blood-brain barrier permeation,” QSAR Comb Sci, vol. 27, no. 5, pp. 586–594, May 2008, doi: 10.1002/QSAR.200710019.

[35] L. Zhang, H. Zhu, T. Oprea, A. Golbraikh, T. I. Oprea, and A. Tropsha, “QSAR modeling of the blood–brain barrier permeability for diverse organic compounds,” Springer, vol. 25, no. 8, pp. 1902–1914, Aug. 2015, doi: 10.1007/s11095-008-9609-0.

[36] S. Kortagere, D. Chekmarev, W. J. Welsh, and S. Ekins, “New predictive models for blood-brain barrier permeability of drug-like molecules,” Pharm Res, vol. 25, no. 8, pp. 1836–1845, Aug. 2008, doi: 10.1007/S11095-008-9584-5.

[37] Z. Gao, Y. Chen, X. Cai, R. X.- Bioinformatics, and undefined 2017, “Predict drug permeability to blood–brain-barrier from clinical phenotypes: drug side effects and drug indications,” academic.oup.com, Accessed: Oct. 01, 2022. [Online]. Available: https://academic.oup.com/bioinformatics/article-abstract/33/6/901/26.

Enhanced Atrial Fibrillation Detection-based Wavelet Scattering Transform with Time Window Selection and Neural Network Integration

Mohamed Elmehdi Ait Bourkha[1*], Anas Hatim[2], Dounia Nasir[3], Said El Beid[4]
Information Technology and Modeling Team (TIM), National School of Applied Sciences (ENSA) of Marrakech,
Cadi Ayyad University (UCA), Marrakech, Morocco[1, 2, 3]
Control and Computing for Smart Systems and Green Energy (CISIEV), ENSA of Marrakech, UCA, Marrakech, Morocco[4]

Abstract—**Atrial Fibrillation (AF), a prevalent anomaly in cardiac rhythm, significantly impacts a substantial portion of the population, with projections indicating an escalation in its prevalence in the near future. This disorder manifests as irregular and accelerated heartbeats originating within the heart's upper chambers known as the atria. Neglecting to address this condition could potentially lead to serious consequences, particularly an elevated susceptibility to stroke and heart failure. This underscores the critical importance of developing an automated approach for detecting AF. In our study, an automatic approach was introduced for classifying short single-lead Electrocardiogram (ECG) recordings signals into four categories: Atrial fibrillation (AF), Normal rhythm (N), Noisy rhythm (~), or Other rhythms (O). The wavelet scattering network (WSN) is employed to extract morphological features from the ECG signals, which are then inputted into an Artificial Neural Network (ANN) with time windows selection and majority vote. The results from the testing data exhibit that our proposed model outperforms the state-of-art models, achieving a remarkable overall accuracy of 87.35% and an F1 score of 89.13%.**

Keywords—Electrocardiogram (ECG); Atrial Fibrillation (AF); Wavelet Scattering Network (WSN); Artificial Neural Network (ANN)

I. Introduction

The ECG is a recording of electrical potential differences on the body surface that result from the electrical activity in the heart [1]. An ECG is produced when a nerve impulse stimulates the heart, causing a current to spread across the body's surface. This current creates a voltage drop ranging from a few microvolts to millivolts, accompanied by variations in the impulse. Typically, these impulses have very low amplitude, necessitating thousands of times of amplification [2]. The ECG is typically a voltmeter that uses up to 12 different leads (electrodes) placed on designated areas of the body [3].

Because of its straightforwardness and non-intrusive characteristics, the ECG has been extensively utilized in the identification of heart diseases [4]. The detection of heart diseases typically involves the analysis of ECG signals, which unveil irregularities commonly known as arrhythmias. These manifestations signify deviations from a regular heart rhythm, potentially causing irregular or abnormal heartbeats, often experienced as palpitations. Arrhythmias are broadly categorized into two main types: ventricular and supraventricular. Atrial Fibrillation (AF), a prevalent condition, falls under the category of supraventricular arrhythmias due to its origination within the heart's upper chambers, the atria. In contrast, ventricular arrhythmias arise from the heart's lower chambers or ventricles. Understanding this differentiation is crucial for accurately identifying and subsequently treating various forms of arrhythmias. This distinction aids medical professionals in precisely classifying the type of arrhythmia observed in a patient, paving the way for more targeted and effective treatment strategies.

Early identification plays a pivotal role in addressing heart arrhythmias, potentially offering significant opportunities to save lives. Utilizing the ECG as a primary diagnostic tool becomes essential in achieving this imperative objective [5]. Nevertheless, the manual interpretation of prolonged ECG recordings introduces a multitude of escalating challenges. As these recordings extend in duration, the intricacies grow, rendering the process more time-consuming, intricate, and arduous. The exhaustive review demanded by these extended recordings not only prolongs the analysis but also heightens the complexity of the interpretation process, making it more challenging [6]. In response to these challenges, cardiologists turn to automated diagnostic algorithms, which streamline the analysis of extensive ECG data [7]. This incorporation of automated methodologies proves to be an invaluable solution, offering a streamlined approach to overcome the hurdles associated with manual interpretation. Consequently, the utilization of these automated tools not only enhances efficiency but significantly improves the precision and management of prolonged ECG recordings. Numerous research initiatives have concentrated on employing classical machine learning models to identify arrhythmias within ECG signals. These models have shown efficacy in analyzing both short-term and long-term ECG readings, primarily focusing on the scrutiny of individual heartbeats within the signal [8-9]. Nevertheless, these models require feature engineering and domain expertise, introducing aspects that are time-consuming and demanding. To address these challenges, deep learning models like Convolutional Neural Networks (CNN) and Long Short-Term Neural Networks have emerged, showcasing impressive performance in detecting arrhythmias [10-11].

AF is the prevailing prolonged cardiac irregularity, found in around 1-2% of the overall population [12-13]. This condition carries notable implications for both mortality and morbidity due to its strong links with various health risks. Individuals with AF face an increased likelihood of experiencing severe outcomes, including stroke, hospitalization, heart failure, and coronary artery disease. The association of AF with these risks underscores its profound impact on cardiovascular health [13-14]. AF's connection to an elevated risk of death further underscores its significance. The irregular heartbeat pattern in AF can lead to blood clots forming within the atria, which might subsequently travel to the brain, causing a stroke. Additionally, the erratic heart rhythm can strain the heart's function over time, potentially culminating in heart failure.

AF impacts more than 12 million people in Europe and North America, and this number is expected to triple in the next 30-50 years. This escalating prevalence underscores the need for increased efforts in diagnosis, treatment, and management to address the growing public health challenge [15]. More notably, AF becomes more common as individuals age. For those between 40-50 years old, the incidence is under 0.5%, while among those 80 and older, it ranges from 5-15%. This age-related rise in AF highlights the need for targeted monitoring and interventions to address the growing risk in the elderly [16]. Recognizing this trend, a multitude of research works and papers have emerged, aiming to develop automatic models utilizing Machine Learning (ML) techniques and Deep Learning (DL) [17]. These endeavors aim to facilitate the diagnosis and early detection of AF, potentially saving the lives of millions across the globe.

In the following sections of this paper, related works were explored, specifically focusing on Automated AF systems from previous endeavors in Section II. Subsequently, our materials and methods cover a data description, data preprocessing, and the features extraction method in Section III. A detailed account was presented to explain the classification model used and the evaluation metrics applied. Further Section IV encompasses the results, where our findings were analyzed and described. The discussion section follows, where our results were compared with existing automated AF detection models. Lastly, the conclusion summarizes our findings, outlines limitations, and suggests avenues for future research in Section V.

II. Related Works

In this paper, a comparison was conducted with state-of-art models designed for AF detection. Notably, Garcia et al. [18] introduced a method that utilizes surface ECG data, capturing variability in ventricular and atrial activities. The approach involves generating time series data from R_R intervals and morphological features of fibrillatory waves in T_Q intervals. The regularity of these time series is quantified using the Coefficient of Sample Entropy (COSEn), and a multi-class Support Vector Machine (SVM) distinguishes between AF, N and O. Their algorithm underwent validation in the PhysioNet Computing in Cardiology Challenge 2017.

Rajpurkar et al. [19] introduced an algorithm surpassing board-certified cardiologist’s proficiency, exhibiting exceptional accuracy in detecting a wide range of heart arrhythmias. The algorithm excels by applying to single-lead wearable monitor electrocardiograms. A sophisticated 34-layer CNN is crucial in mapping ECG sequences to rhythm classes. The study includes a gold standard test set annotated by board-certified cardiologists, serving as a benchmark where the algorithm outperforms individual cardiologists in both recall and precision.

Coppola et al. [20] introduced a data-driven model for automated AF detection from a single ECG lead. The model incorporates features such as heart rate variability, spectral power analysis, and statistical modeling to capture atrial activity nuances. Employing an over-sampling strategy for dataset balance, they crafted a hierarchical classification model predicting ECG signals into AF, N, noise interference or O. Their approach includes a hierarchical bagged ensemble classifier, achieving an average F1 score of 0.7855%.

Makinckas et al. [21] introduced a paradigm utilizing a Long Short Term Memory (LSTM) network, a neural architecture efficiently learning patterns from pre-computed QRS complex features for ECG signal classification. Despite its classification as a deep neural network, their architecture, with a mere 1791 parameters, achieves a remarkable balance between complexity and efficiency. The crux of their methodology lies in the LSTM network's unique ability to comprehend patterns in QRS complex features, facilitating accurate categorization of diverse ECG signals. Their LSTM based model demonstrated effectiveness with a commendable final challenge F1 score of 78% across N, AF, O and ~.

Schwab et al. [22] utilized a richly annotated dataset of 12,186 single-lead ECG recordings. Their approach involved constructing a diverse ensemble of Recurrent Neural Networks (RNN) proficient in discerning differences among N, AF, O, and ~. To enhance temporal learning, they introduced a novel task formulation leveraging ECG signal segmentation into heartbeats, reducing time steps per sequence significantly. Incorporating an attention mechanism further augmented their RNN, enabling the model to focus on specific heartbeats for decision-making. With attention mechanisms, their model achieved an average F1 score of 79%.

Andreotti et al. [23] classified ECG segments into AF, N, O and ~. They conducted a comparative analysis, pitting a feature based classifier against a CNN. Both were meticulously trained on challenge data and augmented with Physionet database. The feature based classifier achieved a 72.0% F1 score during training and 79% on the hidden test set. Meanwhile, the CNN scored 72.1% on the augmented database and 83% on the test set, resulting in a final score of 79%.

Jiménez-Serrano et al. [24] integrated a Feedforward Neural Network (FFNN) for classifying short single-lead ECG segments into N, AF, O and ~. Extracting 72 features from ventricular activity in 8528 ECG records, they conducted a meticulous Feature Selection (FS) process and a detailed grid search for FFNN training parameters. Filtering down to 50 features during FS improved the initial F1 score from 70% to 73%. The FFNN model achieved a final F1 score of 77 %on test data, demonstrating its efficacy in discriminating ECG patterns.

Pandey et al. [25] analyzed ECG data from PhysioNet/CinC Challenge 2017, aiming to differentiate cardiac rhythms, particularly AF, N, O and ~. Their approach combined traditional machine learning with deep neural networks, integrating a Residual Network (ResNet), CNN, Bidirectional LSTM (BiLSTM), and Radial Basis Function (RBF) neural network. The hybrid model achieved an F1 score of 80% and an accuracy of 85% in discerning AF rhythms within the ECG data.

Clifford et al. [26] utilized PhysioNet/CinC Challenge 2017 to distinguish AF from ~, N, or O in short-term ECG recordings. The extensive dataset included 12,186 ECGs, with 8,528 in the public training set and 3,658 in the hidden test set. Utilizing a combination of 45 algorithms, incorporating the LASSO technique, they achieved an F1 score of 86%.

While numerous related works have attempted to detect AF through diverse techniques, including ML, DL, and hybrid models, the prediction accuracy remains notably low. The highest F1 score, achieved by Clifford et al. [26] was 86%, underscores the significance of developing a new approach to enhance the accuracy of AF detection.

III. Materials and Methods

A. Database Description

The research utilizes a publicly available database accessible through this link (https://archive.physionet.org/pn3/challenge/2017/).

This database comprises 8,528 ECG recordings, which were made available as a public training set for utilization in the 2017 PhysioNet/Computing in cardiology challenge, Table I show the distribution of each class label of the dataset [27]. These recordings were captured using an AliveCor handheld device, which automatically uploads the recordings via a mobile phone application. The dataset includes both these recordings and an additional 3,658 recordings retained as a concealed test set. AliveCor provided these recordings for the Challenge. Each ECG recording was sampled at 300 Hz and underwent band-pass filtration through the AliveCor device.

TABLE I. Class Labels Distribution

Class Label	No. Instances
N	5050
AF	738
O	2456
~	284

B. Data Preprocessing

The challenge's dataset draws from a particular database. To ensure a fair basis for comparison with earlier studies that utilized a 20-second interval, each ECG recording was divided into segments. This approach was adopted to establish a uniform time window for analysis, in accordance with previous research practices. Furthermore, the dataset exhibited an imbalance, which could pose training challenges leading to inaccurate classification outcomes. To address this, the Synthetic Minority Over Sampling Technique (SMOTE) was employed [28]. SMOTE generates synthetic instances of the minority class to rebalance the data and enhance classification performance.

Following the application of the SMOTE technique on the 8,528 ECG recordings with segmentations, our dataset was transformed, resulting in 18829 ECG segments. The successful application of SMOTE effectively balanced the class distribution. This balance is pivotal, as it enables the model to achieve precise classification outcomes and effectively differentiate among the four distinct classes. As a consequence of this enhanced discriminatory capacity, our approach will yield elevated accuracy and F1 score results.

In Fig. 1, an ECG segment depicting AF is observed. Notably, the absence of a consistent sinus rhythm is evident through variable R_R intervals. Additionally, the ECG waveform reflects the absence of certain P waves, confirming the presence of AF.

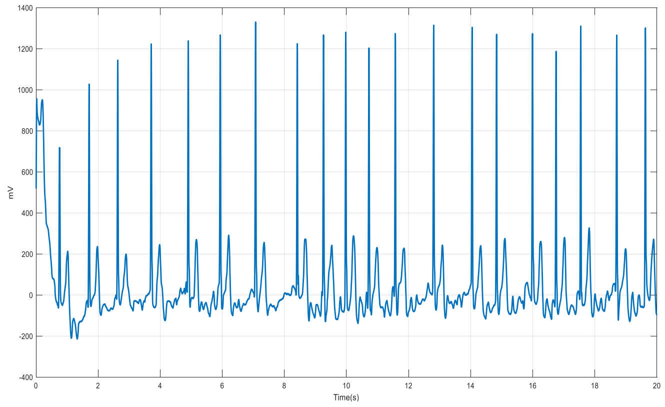


Fig. 1. ECG with atrial fibrillation.

Fig. 2 displays an ECG segment with a normal rhythm. The fixed R_R interval and the presence of all waves and complexes in the ECG waveform are noticeable, indicating the normalcy of the ECG segment.

Fig. 3 reveals an ECG segment with a variable R_R interval, signifying that the ECG segments cannot be classified as N. Notably, the presence of P waves in each ECG heartbeat suggests a deviation from AF. Instead, this ECG segment is categorized as another rhythm.

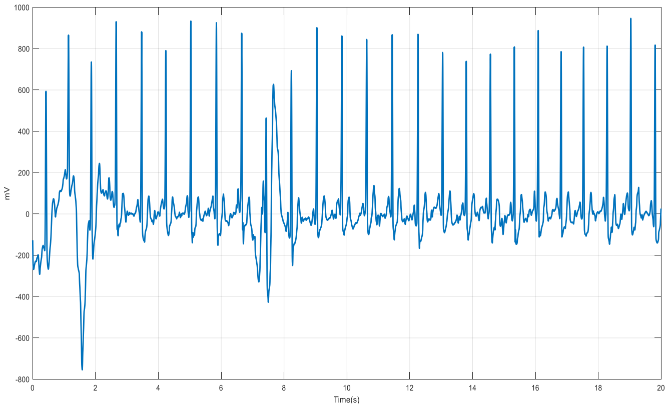


Fig. 2. ECG with normal rhythm.

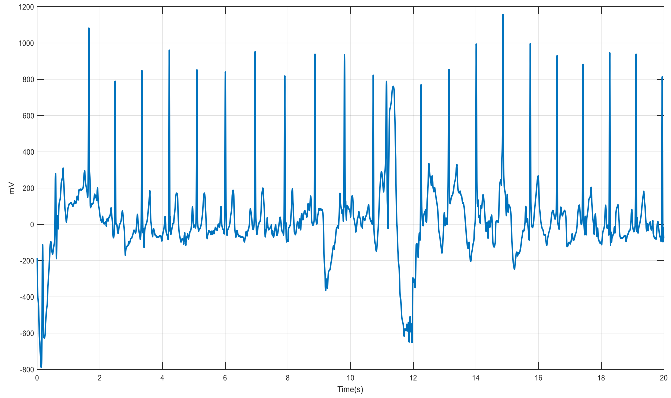


Fig. 3. ECG with other rhythm.

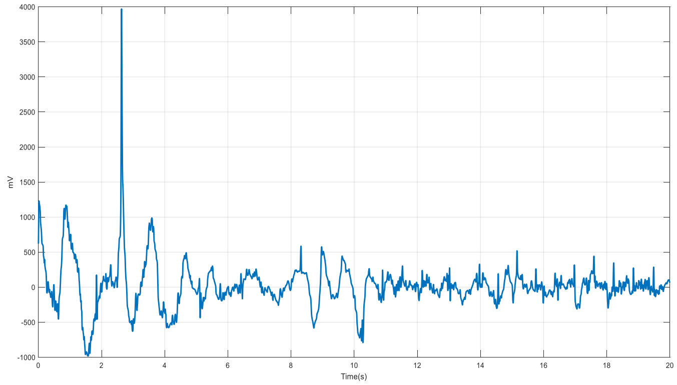


Fig. 4. ECG with noise.

Fig. 4 highlights the absence of identifiable waves or complexes in the ECG segment waveform, indicating that this ECG segment is noisy.

After addressing the imbalance in the dataset through the application of SMOTE, the data was divided into two distinct parts, employing an 80-20% split ratio. This division was carried out with a stratified approach, ensuring that each class label was proportionally represented in both the training and testing sets. Specifically, 80% of the data from each class was allocated to the training set, while the remaining 20% was set aside for testing purposes, which results in a total of 3764 ECG segments for testing and 15065 ECG segments for training. To further enhance the robustness of the model, the training set, constituting 80% of the data, underwent an evaluation process using a five-fold cross-validation technique. This technique involved partitioning the training data into five subsets, training the model on four of these subsets while using the 5th for validation, and then rotating the subsets iteratively. This meticulous combination of techniques facilitated the creation of a well-generalized model capable of effectively addressing the initial data imbalance and yielding reliable performance results.

C. Features Extraction

In this paper, the potential of the WSN was harnessed to meticulously extract an array of intricate morphological characteristics from ECG segments. The wavelet scattering approach is a distinguished member of the deep convolution network family crafted for signal processing, that serve as the linchpin for capturing multifaceted information.

What sets the wavelet scattering method apart is its unique capacity to seamlessly navigate through both the time and frequency domains of signals. By employing wavelets as the foundational building blocks, the network inherently grasps both the rapid fluctuations occurring in short time spans and the nuanced oscillatory patterns occurring over varied frequency ranges. The WSN provides features with translation and rotation invariance, making it suitable for image and audio analysis [29]. It offers stable features for denoising and enables dimensionality reduction for enhanced accuracy.

In this paper, the WSN with Gabor wavelets is utilized due to their morphological similarity to the QRS complex, making them suitable for extracting features from ECG segments [30]. The definition of a Gabor wavelet in Eq. (1) involves the multiplication of a Gaussian function by a complex exponential function. Fig. 5 shows the Gabor wavelet used with its real part, imaginary part, and its low pass filter.

$$\psi(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-t^2}{2\sigma^2}} e^{i\omega t} \tag{1}$$

where, t denotes time, σ represents the standard deviation of the Gaussian function. ω=2πf, where f is the center frequency of ψ and i is the imaginary unit. The Gabor complex wavelet's envelope is a low-pass filter, noted as Φ in Eq. (2).

$$\Phi(t) = |\psi(t)| \tag{2}$$

The scattering network is consisting of three stages as shown in Fig. 6. In the WSN: the 0th order S0, the signal with a low-pass filter Φ was convolved to analyze the slow variations and amplitude in the signal, and provide good time resolution but poor frequency resolution. Moving to the 1st order S_1, a specific-scale wavelet is employed to scrutinize high-frequency components within the ECG segments. And, the 2nd order S2 was proceed to further extract complementary high-frequency components from the analyzed signal. This process enhances our understanding of ECG characteristics.

$$S_0 x(t) = x(t) * \Phi \tag{3}$$

$$S_1 x(t) = |x * \psi_{\sigma_1}| * \Phi \tag{4}$$

$$S_2 x(t) = ||x * \psi_{\sigma_1}| * \psi_{\sigma_2}| * \Phi \tag{5}$$

$$Sx(t) = \{S_0, S_1, \dots, S_n\} \tag{6}$$

After applying the WSN with an invariance scale of 10 seconds and utilizing Q_1=8 and Q_2=1 as the quality factors for the 2 filter banks, a tensor of size 12x205 was obtained for each ECG segment. Fig. 7 shows the frequency bands of first and second filter banks.

In this tensor, columns correspond to the scattering paths within the network, while rows represent time windows. This showcases the WSN's capability to achieve a 59% dimensionality reduction.

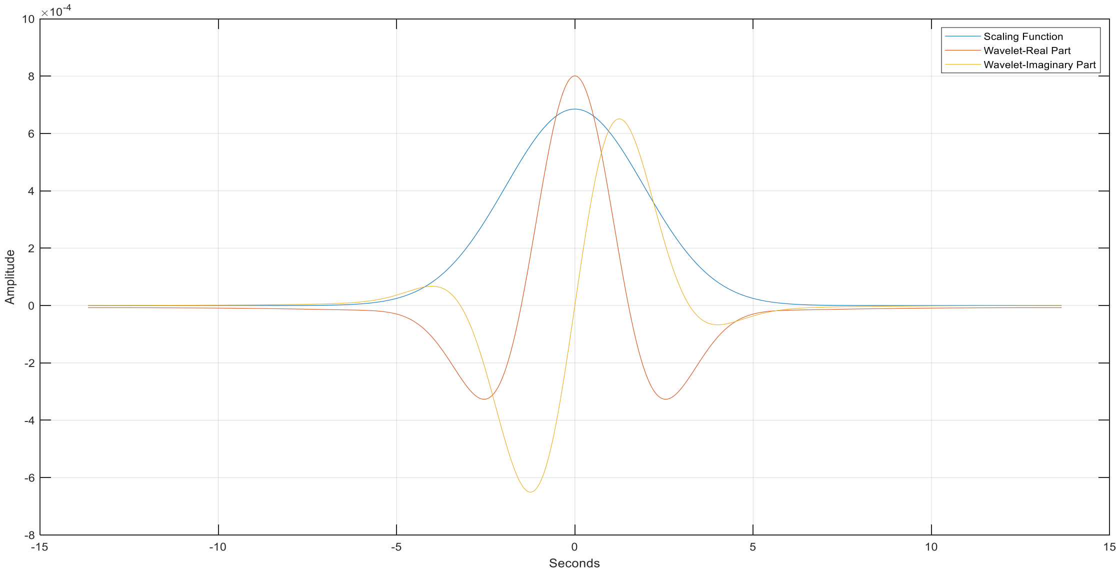


Fig. 5. Gabor complex wavelet.

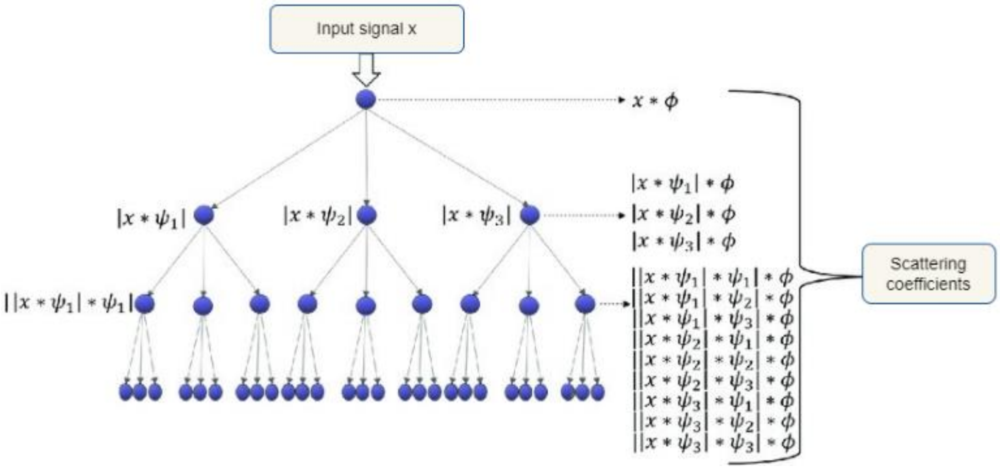

Input signal x
$x * \phi$
$|x * \psi_1|$
$|x * \psi_2|$
$|x * \psi_3|$
$|x * \psi_1| * \phi$
$|x * \psi_2| * \phi$
$|x * \psi_3| * \phi$
$||x * \psi_1| * \psi_1|$
$||x * \psi_1| * \psi_1| * \phi$
$||x * \psi_1| * \psi_2| * \phi$
$||x * \psi_1| * \psi_3| * \phi$
$||x * \psi_2| * \psi_1| * \phi$
$||x * \psi_2| * \psi_2| * \phi$
$||x * \psi_2| * \psi_3| * \phi$
$||x * \psi_3| * \psi_1| * \phi$
$||x * \psi_3| * \psi_2| * \phi$
$||x * \psi_3| * \psi_3| * \phi$
Scattering coefficients

Fig. 6. Scattering network.

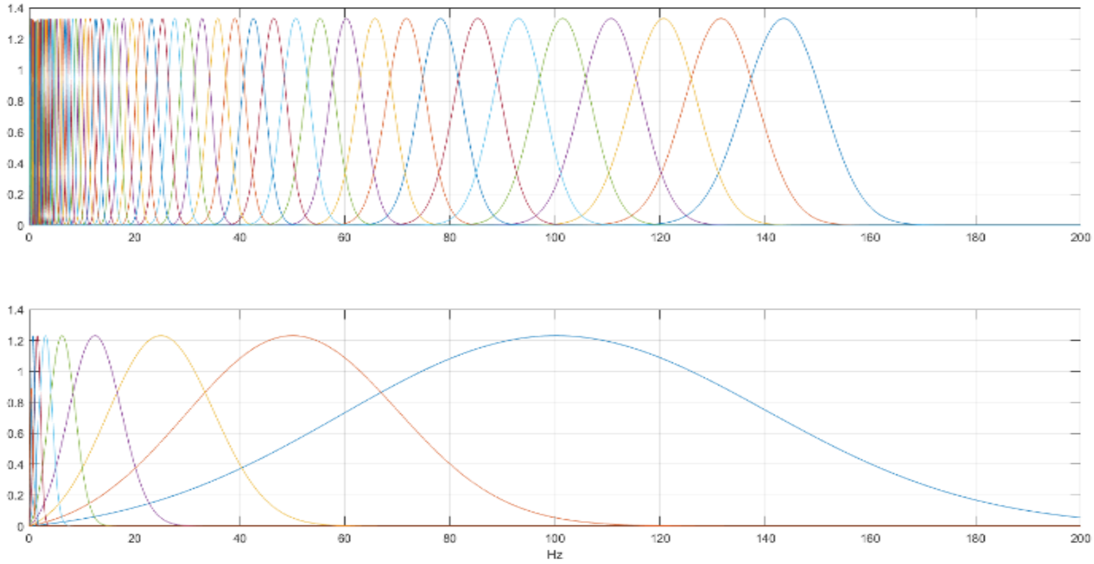


Fig. 7. Frequency band of the first and the second filter bank.

Consequently, the training dataset becomes 15065x12x205, and the testing dataset becomes 3764x12x205 in tensor dimensions. Subsequently, reshaping the training and testing datasets into an appropriate format for classifiers result in feature matrices of size 180780x205 and 45168x205 for training and testing, respectively.

The scalogram coefficients depicted in Fig. 8 showcase the outcomes of convolving the AF in ECG segment, as depicted in Fig. 1, with the real and imaginary components of Gabor wavelets within the initial filter bank.

This visual representation is exceptional in its ability to delineate the various frequencies present within the signal while associating each frequency with its respective temporal occurrence.

Moreover, the convolution process with these filters allows for the computation of similarity between the signal and the wavelets. These features are instrumental in providing insights into the amplitudes and frequencies within the signal, which in turn play a crucial role in accurately predicting atrial fibrillation.

The preceding WSN was implemented in the MATLAB environment, with an invariance scale set to 10 seconds and a sampling frequency of 300 Hz utilized.

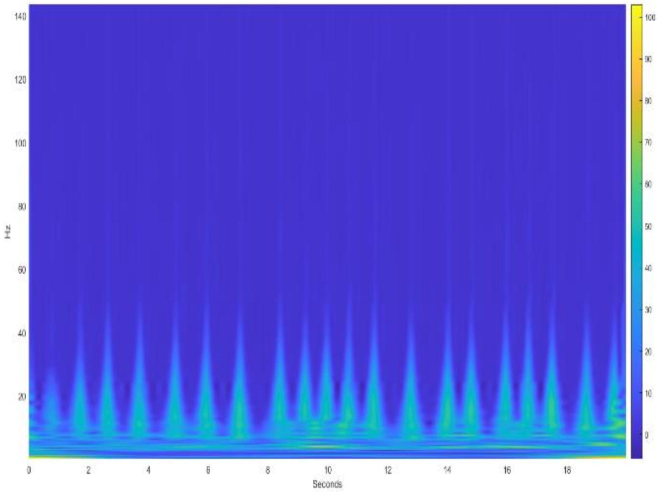


Fig. 8. Scalogram coefficients for the first filter bank.

D. Classification Model

In our study, an ANN was constructed with a single hidden layer containing 200 neurons. The goal was to effectively differentiate among classes A, N, O, and ~. The ReLU was used as activation function layer and Softmax as output activation layer. To enhance the model's performance, the Limited-memory-Broyden–Fletcher–Goldfarb–Shanno (LBFGS) solver was adopted, setting the maximum number of iterations to 1000. Throughout our analysis, we focused on refining the accuracy and efficacy of our approach.

The first hidden layer equation for each neuron j can be described as follows:

$$Z_j = \sum_{i=1}^{n} W_{ij}\, X_i + \; b_j \tag{7}$$

$$a_j = \max(0, Z_j) \tag{8}$$

where, n is the number of input features, W_{ij} is the weight between input features i and hidden neuron j, b_j is the bias term of hidden neurone j, and a_j is the output of neuron j after applying the ReLU activation function.

For each class c of the output layer:

$$Z_c = \sum_{j=1}^{200} W_{jc}\, a_j + \; b_c \tag{9}$$

$$Y_c = \frac{e^{Z_c}}{\sum_{k=1}^{4} e^{Z_K}} \tag{10}$$

where, Y_c is the predicted probability for class c after applying the Softmax activation function.

The ANN was constructed within the MATLAB environment, with initial weights determined using the 'glorot' method and initial biases set to zeros. The maximum number of iterations was set at 1000, and the gradient tolerance was established as 10^{-6}. The lambda parameter remained fixed at 0.

E. Evaluation Metrics

Our classifier model's performance was evaluated using metrics like accuracy, precision, recall, specificity, and the F1 score. Accuracy measures correct classifications, precision gauges accurate positive predictions, recall assesses positive instances captured, specificity measures accurate negative predictions, and the F1 score balances precision and recall.

Accuracy: It evaluates the ratio of accurate forecasts generated by the model among all the predictions it makes.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{11}$$

Precision: It evaluates how well the model correctly identifies positive cases, indicating the ratio of true positives to all positive predictions.

$$\text{Precision or PPV} = \frac{\text{TP}}{\text{TP+FP}} \tag{12}$$

Sensitivity: It measures the percentage of real positive instances accurately detected by the model and also referred as recall or the true positive rate.

$$\text{Recall or Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{13}$$

Specificity: It evaluates the model's capability to correctly recognize negative cases through the measurement of true negative prediction’s proportion.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{14}$$

F1 Score: It calculates a balanced evaluation of the model's performance by taking the harmonic mean of both precision and recall, merging them into a single metric.

$$\text{F1 score} = \frac{2*\text{precision}*\text{sensitivity}}{\text{precision+sensitivity}} \tag{15}$$

F. System Description

The implementation of all algorithms was carried out using MATLAB version R-2021b on a Windows server. The system employed for execution featured an Intel (R), Core (TM), i5, CPU 6300U, processor clocked at 2.40 GHz, with a 12 GB RAM capacity and operating on a 64 bits architecture.

IV. Results and Discussion

A. Results

The objective of our paper was to classify ECG signals into four classes: AF, N, O, and ~. Numerous ML models were tested to assess their effectiveness in distinguishing these classes. Our findings, as indicated in Table II, revealed that an ANN with a single hidden layer and a size of 100 neurons yielded the best classification results. This architecture achieved an accuracy of 87.2% on the validation data and 81.1% on the testing data.

Interestingly, our observations also highlighted that increasing the size of the first hidden layer led to improved accuracy in both training and testing data. To delve deeper into this phenomenon, the impact of various hidden layer sizes was examined on classification outcomes. The results, depicted in Fig. 9, confirmed that accuracy consistently improved with larger layer sizes. However, it's important to note that excessively increasing the layer size can result in heightened computational costs and reduced prediction speed. Therefore, the layer size was adjusted at 200 neurons, which balanced accuracy and computational efficiency. With this configuration, a remarkable accuracy rates were achieved: 90.35% on the validation data and 82.10% on the testing data.

Upon applying the WSN to each ECG signal, a tensor of dimensions 12x205 was obtained. After reshaping the data for training and testing, an ANN was employed with a hidden layer size of 200. The ANN produced 12 results corresponding to different time windows. Consequently, each time window exhibited distinct validation and testing accuracies.

In Fig. 10, the impact of these time windows on validation and testing accuracy was analyzed. The results revealed that the 5th time window yielded the highest classification performance, achieving 91.85% accuracy on validation data and 83.95% accuracy on testing data.

This proposed approach, known as Time Window Selection showcased a noteworthy enhancement in accuracy. Validation accuracy improved from 90.35% to 91.85%, while testing accuracy saw an increase from 82.10% to 83.95%.

Enhancing the testing accuracy is achievable through the strategic selection of optimal time windows for classification, followed by applying a majority vote approach. Fig. 10 illustrates that starting from the 3rd time window, there is a noticeable improvement in validation accuracy.

To harness this insight, the results from the ANN with a 200-layer size across 12-time windows were utilized, focusing on the 3rd through the 10th time windows, which displayed superior validation accuracy. Employing a majority vote technique on these eight selected time windows, we observed a significant enhancement in testing accuracy and F1 score.

TABLE II. Performance Comparaison of Different Machine Learning Models

Models / Accuracy	Decision Tree	Narrow Neural Network layer size: 10	Medium Neural Network layer size: 25	Wide Neural Network layer size: 100	Bilayered Neural Network First layer size: 10 Second layer size: 10
Validation Data %	61.8	75.2	80.2	87.2	75.8
Testing Data %	62.1	74.4	78.0	81.1	72.1

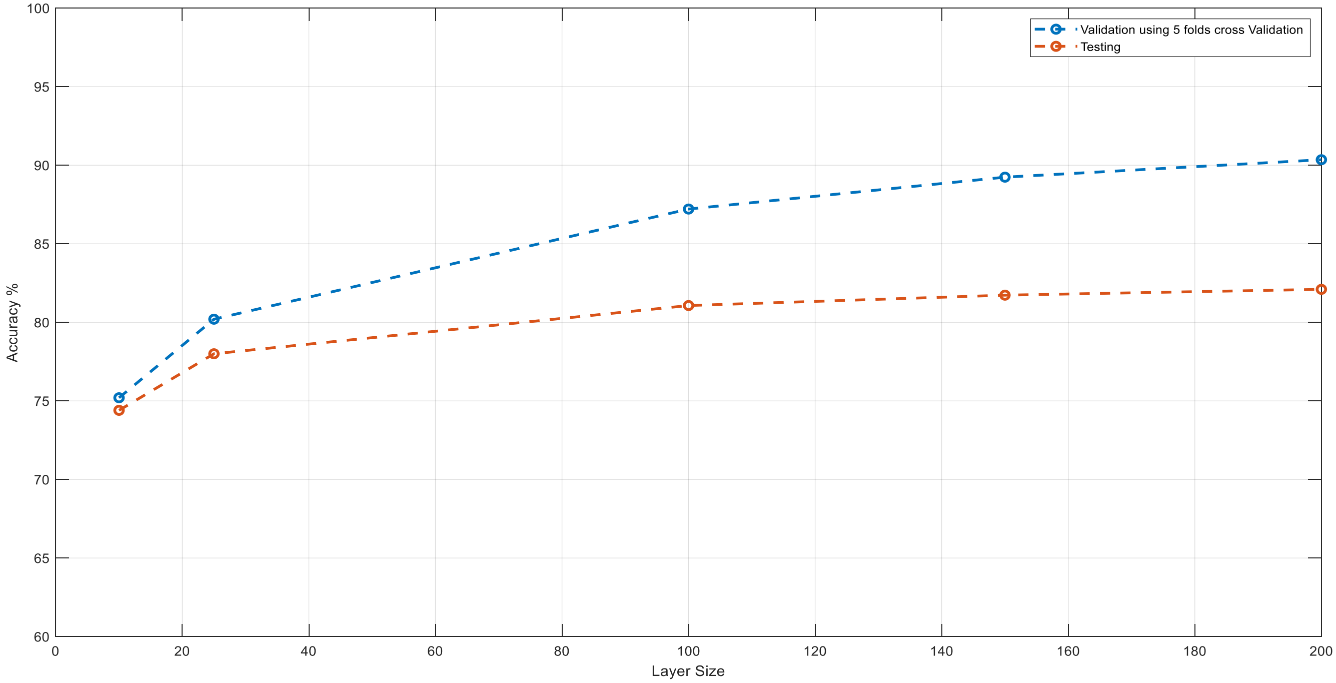


Fig. 9. ANN layer and its impact on validation and testing accuracy.

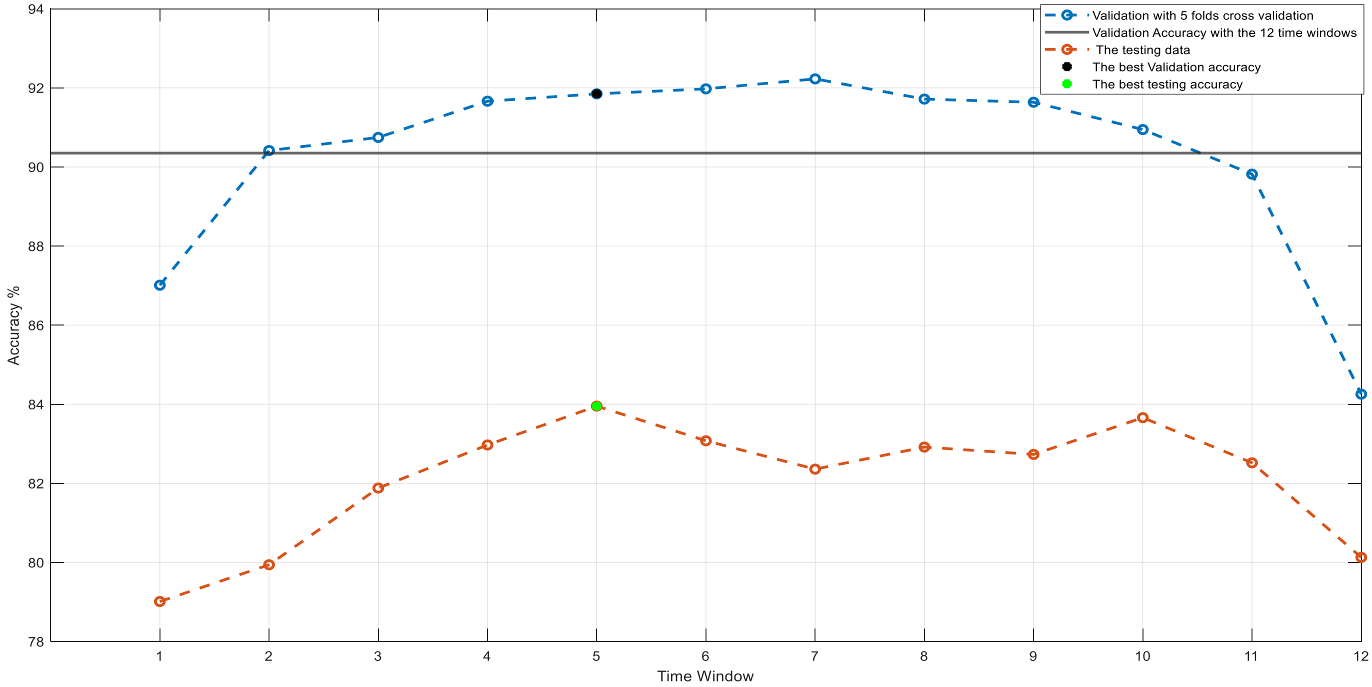


Fig. 10. Time windows impact on validation and testing accuracies.

TABLE III. TESTING RESULTS USING WSN + ANN + TIME WINDOWS SELECTION + MAJORITY VOTE

Class Name	Precision %	Recall %	Specificity %	F1 score %
AF	89.05	93.23	98.11	91.09
N	85.19	89.37	89.86	87.23
O	84.19	76.73	93.48	80.29
~	97.40	98.43	99.53	97.91
Average	88.96	89.44	95.24	89.13

TABLE IV. SUMMERIZED RESULTS ON THE VALIDATION AND THE TESTING DATASETS

	Methodology	**Accuracy %**	**F1 score %**
5 Folds Cross Validation on the Training Dataset	***WSN + ANN with 12 Time window***	90.35	91.53
	WSN + ANN + 5th Time window	91.85	93.02
Testing Dataset	***WSN + ANN with 12 Time window***	82.10	84.22
	WSN + ANN + 5th Time window	83.95	85.94
	WSN + ANN + Time Windows selection + Majority Vote	87.35	89.13

This approach led to a testing accuracy of 87.35% and an F1 score of 89.13%.

Detailed testing results are available in Table III, while Table IV provides a concise summary of the outcomes obtained through various methodologies.

Fig. 11 displays the various steps outlined in this paper for classifying ECG signals as N, AF, ~, or O.

B. Discussion

In this study, we compared the outcomes produced by our classification method with those of previously established state-of-the-art models. Our model, which combines WSN, ANN, Time window, and Majority vote technique, achieves the highest overall accuracy. When contrasted with the findings of Pandey et al. [18], as presented in Table V, our model outperforms the state of art models in term of accuracy.

A comparative analysis of our results with those of other studies was conducted, specifically focusing on the F1 score.

The highest F1 score, 89.13%, was attained in our paper. Nonetheless, as shown in Table VI, it remains evident that our proposed methodology surpasses the performance of preceding works in terms of F1 score.

An extensive examination of the model's complexity was conducted, encompassing the neuron count in the ANN and the overall count of learnable parameters. Additionally, the time taken for feature extraction from a single ECG segment, quantified at approximately 67.3 milliseconds as displayed in Table VII. The predictive speed of our ANN successfully attains a rate of 22,584 ECG segments per second, showcasing a remarkably high prediction speed for detecting atrial fibrillation within ECG segments.

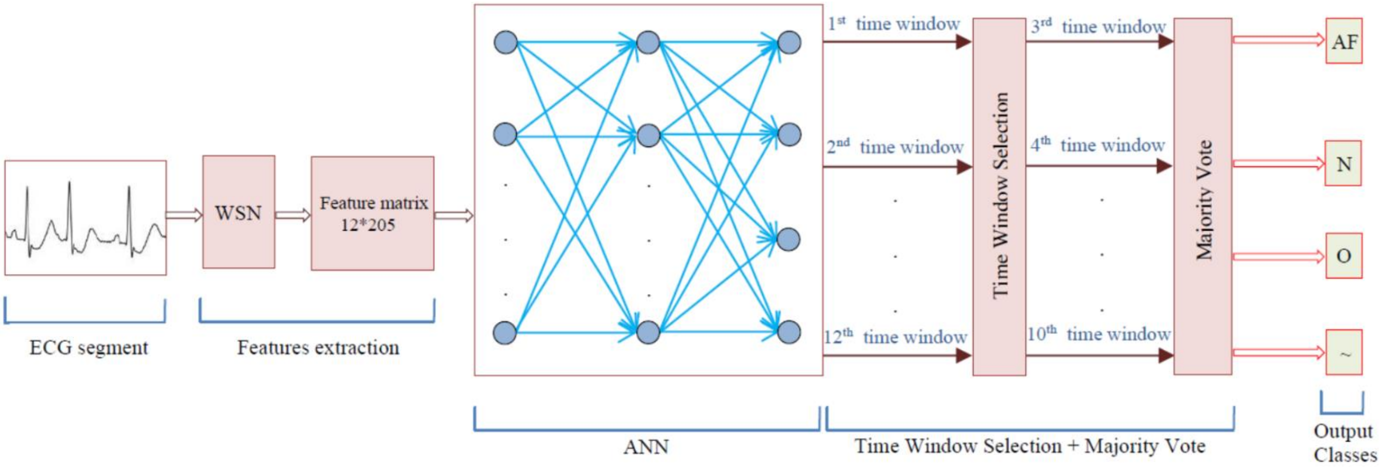


Fig. 11. ECG segments classification process.

TABLE V. OVERALL ACCURACY COMPARAISON WITH OTHER PREVIOUS WORK

Study	Methodology	Accuracy %
Pandey et al. [25]	***ResNet***	84.40
	ResNet + LSTM	82.87
	ResNet + RBF	84.56
Present Study	***WSN + ANN***	82.10
	WSN + ANN + 5th Time window	83.95
	WSN + ANN + Time windows selection + Majority Vote	**87.35**

TABLE VI. F1 SCORE COMPARAISON WITH OTHER PREVIOUS WORKS

Study	F1 score %
García et al. [18] (2017)	73
Rajpurkar et al. [19] (2017)	79.9
Coppola et al. [20] (2017)	78.55
Maknickas et al. [21] (2017)	78
Schwab et al. [22] (2017)	79
Jimenez-Serrano et al. [23] (2017)	77
Andreotti et al. [24] (2017)	79
Clifford et al. [26] (2017)	86.8
Pandey et al. [25] (2022)	80.56
Present Work (2023)	**89.13**

TABLE VII. COMPLEXITY ANALYSIS

No. Neurons	204
No. Learnabeles	42004
Feature Extraction Time for 1 ECG segment	67.3 millisecond
Prediction Speed of ANN	22584 ECG segment/s
Training Time	240.7 minutes

V. CONCLUSION

In our research, an innovative approach for the automated classification of ECG signals and the detection of atrial fibrillation was presented.

Our technique leverages a combination of WSN with ANN, Time Windows Selection, and Majority Vote to yield promising results when compared to prior studies, achieving an accuracy of 87.35%, a precision of 88.96%, a recall of 89.44%, a specificity of 95.24%, and an F1 score of 89.13%.

Although, our proposed approach has shown a good performance, it still has some limitations. Firstly, the ANN performance was dependent on the accuracy and reliability of the features derived from the raw ECG data before being inputted. Secondly, the current method is challenged by a significant computational burden due to the feature extraction process.

In forthcoming work, we intend to explore a technique that mitigate the computational costs associated with our proposed model

To addressing these identified constraints, future endeavors will focus on enhancing the proposed model through the application of dimensionality reduction techniques utilizing machine learning. This enhancement aims to streamline the feature space, thereby lowering the computational load in the classification process without compromising the ANN efficacy.

DATA AVAILABILITY

ECG readings were taken from: https://archive.physionet.org/pn3/challenge/2017/.

CONFLECTS OF INTEREST

We confirm that all authors declare no conflicts of interest.

ACKNOWLEDGMENT

This work was supported by the National Scientific and Technical Research Centre (CNRST), Morocco.

REFERENCES

[1] J. Sundnes, G. T. Lines, X. Cai, B. F. Nielsen, K. A. Mardal, A. Tveito, “Computing the electrical activity in the heart,” Springer Science & Business Media, (Vol. 1), 2007.

[2] M. K. Islam, G. Tangim, T. Ahammad, M. R. H. Khondokar, “Study and analysis of ecg signal using matlab &labview as effective tools,” International journal of Computer and Electrical engineering, *4*(3), 404. 2012.

[3] A. K. M. F. Haque, M. H. Ali, M. A. Kiber, M. T. Hasan, “Detection of small variations of ECG features using wavelet,” ARPN Journal of Engineering and applied Sciences, 4(6), 27-30, 2009.

[4] E. J. D. S. Luz, W. R. Schwartz, G. Cámara-Chávez, D. Menotti, “ECG-based heartbeat classification for arrhythmia detection: A survey,” Computer methods and programs in biomedicine, 127, 144-164, 2016.

[5] A. U. Rahman, R. N. Asif, K. Sultan, S. A. Alsaif, S. Abbas, M. A. Khan, A. Mosavi, “ECG classification for detecting ECG arrhythmia empowered with deep learning approaches,” Computational intelligence and neuroscience, 2022.

[6] S. S. Hussain, F. Noman, H. Hussain, C. M. Ting, S. R. G. S. bin Hamid, H. Sh-Hussain, ..., J. Ali, “A Brief Review of Computation Techniques for ECG Signal Analysis,” In Proceedings of the Third International Conference on Trends in Computational and Cognitive Engineering: TCCE 2021 (pp. 223-234). Singapore: Springer Nature Singapore, February 2022.

[7] L. Saclova, A. Nemcova, R. Smisek, L. Smital, M. Vitek, M. Ronzhina, “Reliable P wave detection in pathological ECG signals,” Scientific Reports, 12(1), 6589, 2022.

[8] R. Holgado-Cuadrado, C. Plaza-Seco, L. Lovisolo, M. Blanco-Velasco, “Characterization of noise in long-term ECG monitoring with machine learning based on clinical criteria,” Medical & Biological Engineering & Computing, 1-14, 2023.

[9] X. Dong, W. Si, “Heartbeat dynamics: a novel efficient interpretable feature for arrhythmias classification,” IEEE Access, 2023.

[10] V. Rawal, P. Prajapati, A. Darji, “Hardware implementation of 1D-CNN architecture for ECG arrhythmia classification,” Biomedical Signal Processing and Control, 85, 104865, 2023.

[11] M. Karri, C. S. R. Annavarapu, “A real-time embedded system to detect QRS-complex and arrhythmia classification using LSTM through hybridized features,” Expert Systems with Applications, 214, 119221, 2023.

[12] G.Y.H. Lip, L. Fauchier, S.B. Freedman, I. Van Gelder, A. Natale, C. Gianni, S. Nattel, T. Potpara, M. Rienstra, H. Tse, D.A. Lane, “Atrial fibrillation,” Nature Reviews Disease Primers 2, 16016, 2016.

[13] Developed with the Special Contribution of the European Heart Rhythm Association (EHRA), Endorsed by the European Association for Cardio-Thoracic Surgery (EACTS), Authors/Task Force Members, Camm, A. J., Kirchhof, P., Lip, G. Y., ... & Zupan, I. (2010). Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). European heart journal, 31(19), 2369-2429.

[14] R. Colloca, “Implementation and testing of atrial fibrillation detectors for a mobile phone application, 2013.

[15] I. Savelieva, J. Camm, J., “Update on atrial fibrillation: part I,” Clinical Cardiology: An International Indexed and Peer-Reviewed Journal for Advances in the Treatment of Cardiovascular Disease, 31(2), 55-62, 2008.

[16] G. V. Naccarelli, H. Varker, J. Lin, K. L. Schulman, “Increasing prevalence of atrial fibrillation and flutter in the United States,” The American journal of cardiology, 104(11), 1534-1539, 2009.

[17] A. Ghrissi, D. Almonfrey, F., Squara, J. Montagnat, V. Zarzoso, “Identification of spatiotemporal dispersion electrograms in atrial fibrillation ablation using machine learning: A comparative study,” Biomedical Signal Processing and Control, 72, 103269, 2022.

[18] M. García, J. Ródenas, R. Alcaraz, J. J. Rieta, “Atrial fibrillation screening through combined timing features of short single-lead electrocardiograms,” In 2017 Computing in Cardiology (CinC) (pp. 1-4). IEEE, September 2017.

[19] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, A. Y. Ng, “Cardiologist-level arrhythmia detection with convolutional neural networks,” arXiv preprint arXiv:1707.01836, 2017.

[20] E. E. Coppola, P. K. Gyawali, N. Vanjara, D. Giaime, L. Wang “Atrial fibrillation classification from a short single lead ECG recording using hierarchical classifier,” In 2017 Computing in Cardiology (CinC) (pp. 1-4). IEEE, September 2017.

[21] V. Maknickas, A. Maknickas, “Atrial fibrillation classification using qrs complex features and lstm,” In 2017 Computing in Cardiology (CinC) (pp. 1-4). IEEE, September 2017.

[22] P. Schwab, G. C. Scebba, J. Zhang, M. Delai, W. Karlen, “Beat by beat: Classifying cardiac arrhythmias with recurrent neural networks,” In 2017 Computing in Cardiology (CinC) (pp. 1-4). IEEE, September 2017.

[23] F. Andreotti, O. Carr, M. A. Pimentel, A. Mahdi, M. De Vos, “Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG,” In 2017 Computing in Cardiology (CinC) (pp. 1-4). IEEE, September 2017.

[24] S. Jiménez-Serrano, J. Yagüe-Mayans, E. Simarro-Mondéjar, C. J. Calvo, F. Castells, J. Millet, “Atrial fibrillation detection using feedforward neural networks and automatically extracted signal features,” In 2017 Computing in Cardiology (CinC) (pp. 1-4). IEEE, September 2017.

[25] S. K. Pandey, G. Kumar, S. Shukla, A. Kumar, K. U. Singh, S. Mahato, “Automatic Detection of Atrial Fibrillation from ECG Signal Using Hybrid Deep Learning Techniques,” Journal of Sensors, 2022.

[26] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, ..., R. G. Mark, “AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017,” In 2017 Computing in Cardiology (CinC) (pp. 1-4). IEEE. F1 score :0.868, September 2017.

[27] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. Ivanov, R. G. Mark, ..., H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” circulation, 101(23), e215-e220, 2000.

[28] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P0 Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” Journal of artificial intelligence research, 16, 321-357, 2002.

[29] J. Bruna, S. Mallat, « Invariant scattering convolution networks,” IEEE transactions on pattern analysis and machine intelligence, 35(8), 1872-1886, 2013.

[30] S. S. Goh, A. Ron, Z. Shen, “Gabor and wavelet frames,” World Scientific, (Vol. 10), 2007.

Advancing Road Safety: Precision Driver Detection System with Integrated Overspeed, Alcohol Detection, and Tracking Capabilities

Jamil Abedalrahim Jamil Alsayaydeh[1*], Mohd Faizal bin Yusof[2],
Khivisha S.Mohan[3], A K M Zakir Hossain[4], Serhii Leoshchenko[5]
Department of Engineering Technology-Fakulti Teknologi and Kejuruteraan Elektronik and Komputer (FTKEK),
Universiti Teknikal Malaysia Melaka (UTeM), 76100 Melaka, Malaysia[1, 3, 4]
Research Section-Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates[2]
Department of Software Tools, National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine[5]

Abstract*—In response to ongoing concerns about road accidents linked to overspeeding and drunk driving, this study introduces a groundbreaking solution: The Integrated Driver Safety system. It is a comprehensive vehicle safety system designed for real-time prevention. Crafted with cutting-edge components including ESP32, MQ3 sensor, relay, and GPS, this system operates on a dual framework. It swiftly detects instances of over speeding, triggering immediate email alerts, while concurrently inhibiting engine ignition upon detecting alcohol consumption, actively thwarting drunk driving attempts. This proactive approach not only provides real-time notifications but physically prevents intoxicated driving, drastically reducing accidents caused by these factors. With an impressive overspeed detection accuracy surpassing 95% and an efficient alcohol monitoring system, this technology cultivates responsible driving habits. Its potential widespread adoption foretells a future where road safety reaches unprecedented levels, underscoring the industry's dedication to innovation and safer driving experiences. Through this research, a compelling case emerges for the global embrace of these innovative preventive measures, illuminating a path toward significantly enhanced road safety standards.

Keywords*—Integrated driver safety; overspeed detection system; alcohol monitoring technology; comprehensive vehicle security; real-time accident prevention; ESP32 GPS safety; responsible driving solutions

I. Introduction

In the global landscape, road safety remains a paramount concern, with traffic accidents claiming millions of lives annually. The alarming rise in traffic accidents is intrinsically linked to two major factors: over speeding and drunk driving. This paper critically explores the nexus between these perilous behaviors and their devastating consequences on road safety. Through a meticulous analysis of existing literature and statistical data, this study sheds light on the urgent need for comprehensive measures to combat this pressing issue.

The Malaysian Road Transport Department's data for the year 2019 starkly reveals the gravity of the situation, with 4,715 fatal road accidents attributed to over speeding and drunk driving. Malaysia, recognizing the dire implications of these behaviors, has implemented stringent laws to curb them. Operating a vehicle while intoxicated, a dangerous behavior that significantly jeopardizes public safety, is met with severe legal repercussions, such as arrests, criminal charges, and fines. Despite these strict measures, there are individuals who continue to drive under the influence and exceed speed limits, resulting in serious accidents like head-on collisions and rear-end crashes. Acknowledging these challenges, the government has escalated its initiatives by increasing police presence, imposing fines, and running public awareness campaigns. However, relying solely on conventional methods is not enough to address this issue comprehensively.

This research explores the creation and application of a cutting-edge solution: The Internet of Things (IoT)-based car safety system known as the Overspeed and Alcohol Detection System with Tracking (OADS-T). Unlike traditional methods reliant on post-incident investigations, the system's advanced sensors and GPS technology enable instantaneous detection of alcohol consumption and instances of speeding, ensuring swift notifications to authorities for prompt interventions. This addresses the critical need for timely responses to prevent accidents and reduce emergency response times. Moreover, this system facilitates comprehensive data collection by integrating precise GPS tracking and sophisticated algorithms for overspeeding and alcohol level detection. This comprehensive approach surpasses the limitations of conventional methods, providing nuanced insights into diverse driving behaviors and road conditions. The system's ability to gather detailed information contributes to a more thorough understanding of the factors contributing to road safety challenges.

The main objective of this article is to create and confirm the effectiveness of the OADS-T system in improving road safety. By designing advanced algorithms for overspeeding and alcohol level detection and incorporating precise GPS tracking, this research aims to tackle the complexities of implementation. Investigating varied vehicles and diverse road conditions in Malaysia, this study explores potential challenges and benefits, aiming to save lives, curtail accidents, and reduce the economic and societal costs associated with reckless driving.

To substantiate the claims made in this introduction, reference is made to the study by [1], which underscores the critical need to evaluate existing interventions in overspeed detection systems. The intricately constructed systems, integrating cutting-edge technologies such as radar, GPS, and others, form the foundation of this research.

Beyond its immediate impact, the successful implementation of the OADS-T system could pioneer advancements in global car safety technology, contributing to a safer transportation environment worldwide [2] [3]. This research project signifies a crucial step toward investigating cutting-edge technologies' potential to prevent irresponsible driving behaviors, ultimately making a substantial contribution to the field of road safety.

The remainder sections of this paper will delve into the study's background and related works in Section II, followed by a detailed exploration of system implementation and testing in Section III. Section IV will present the results and analysis and finally, the conclusion is described in Section V.

II. Background of the Study

In the realm of road safety, preventing accidents caused by over speeding and drunk driving has become an urgent global priority. This research paper conducts a meticulous examination of existing vehicle safety features designed to address these perilous behaviors. Numerous studies and projects in the field have demonstrated the efficacy of various frameworks, leading to a deeper understanding of the challenges and potential solutions.

A. *Comprehensive Analysis of Overspeed Detection Systems*

In the realm of transportation safety, overspeed detection systems have emerged as indispensable tools, mitigating the risks entangled with high-speed driving. This study delves into a profound analysis, aiming to unearth the impact of performance and speed on the heightened vulnerability of automobiles to crashes. By meticulously analyzing diverse data sources and conducting an extensive literature review, this paper embarks on a journey to explore the significant risks posed by negligent driving behaviors, shedding light on the urgent need for comprehensive interventions.

Despite the implementation of speed limits and protective measures, fatal crashes persist unabated, emphasizing the criticality of evaluating the effectiveness of existing interventions [1]. This paper seeks to dissect overspeed detection systems and their pivotal role in informing drivers or initiating corrective action when a vehicle surpasses predetermined speed limits. These systems, intricately constructed through the integration of various cutting-edge technologies such as radar, GPS, and other similar advancements, form the crux of this discussion.

In a pivotal study, a dedicated research team meticulously examined the efficacy of a GPS-based speeding detection system, ingeniously employing an in-vehicle display to alert drivers of their excessive speed [4]. This study heralds a notable advancement in the transportation landscape—a potential automated technology capable of adjusting vehicle speed to adhere to designated speed limits, if deemed necessary. The findings of this study echo a resounding success, effectively mitigating instances of speeding and ushering in significant enhancements in overall road safety. The implementation of software harnessing GPS or RFID technology to generate geographic limit features is complemented by a robust set of geofencing capabilities. These geofences, intangible boundaries within a digital environment, serve as the linchpin in surveillance and location tracking systems, based on global satellite navigation services. The device model, a testament to technological innovation, incorporates Arduino MEGA, GSM, GPS, and Flex sensors, with a Neo 6m GPS module and GSM module (sim800l v2.0). The choice of Arduino Mega stands out for its affordability, versatility, user-friendliness, and extensive community support. The integration is streamlined with four serial ports, simplifying sensor interfaces.

Another ground-breaking study [5] delved into radar-based overspeed detection systems, specifically in sensitive areas such as school zones. The study's findings were resounding—these systems significantly reduced speeding incidents. Remarkably, the mere presence of these systems had a deterrent effect, compelling drivers to decelerate even in the absence of active detection and correction.

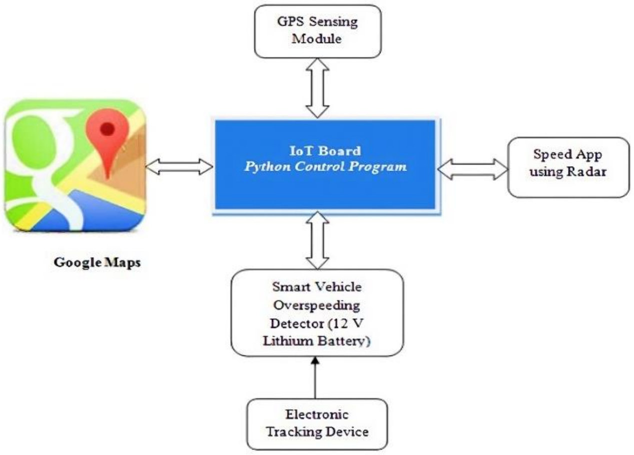

GPS Sensing Module
IoT Board
Python Control Program
Speed App using Radar
Google Maps
Smart Vehicle Overspeeding Detector (12 V Lithium Battery)
Electronic Tracking Device

Fig. 1. Illustrates an innovative IoT-based smart vehicle speeding detection system [6].

This study examines an innovative IoT-based smart vehicle speeding detection system, illustrated in Fig. 1, as detailed in source [6]. A smart car over-speeding sensor was used in conjunction with IoT to reduce the vehicle's speed in certain locations such as susceptible to accident zones. Disasters can be avoided if this smart sensor technology is utilized to set safety criteria. The data is remotely sent by the system. If the sensor detects an over-speeding vehicle, an alarm is triggered.

In recent years, considerable efforts have been made to develop overspeed detection systems utilizing various technologies such as video cameras, machine learning algorithms, radar, and GPS. The efficacy of certain systems can be observed. In a scholarly investigation, the utilization of video cameras and machine learning methodologies was examined as a means to categorize vehicles based on their velocity [7]. The present study reveals that the system under investigation exhibits a high level of accuracy in its ability to identify and classify automobiles, thereby suggesting its potential utility in the realm of detecting and enforcing instances of speeding violations. Studies have demonstrated that these technologies have the potential to be beneficial in lowering the number of incidents caused by speeding in any one of these environments.

The theoretical underpinning of overspeed detection systems revolves around the profound impact of speed on traffic accidents. Excessive speed stands as a leading contributor to road fatalities, necessitating robust measures to curtail speeding incidents. Technologies such as radar, GPS, and machine learning play pivotal roles in enhancing road safety. They enable precise speed monitoring and deliver timely driver warnings, empowering individuals to adjust their driving behavior proactively. As we move forward, it is imperative to incorporate these findings into the development of overspeed detection systems. By leveraging appropriate technologies and integrating them intelligently, we can effectively detect and mitigate overspending, ushering in a new era of enhanced road safety. Through continuous innovation and strategic implementation, these advanced systems stand poised to significantly contribute to the global endeavor of reducing road accidents and saving lives on our highways.

B. *Advancing Road Safety Through Alcohol Detection Systems*

Alcohol detection systems stand as pivotal tools in curbing drunk driving and mitigating alcohol-related accidents. These systems employ sophisticated technologies, including breathalyzers, to measure a driver's breath alcohol content and prevent vehicle ignition if the limit is exceeded.

Numerous studies have underscored the efficacy of alcohol detection systems in enhancing road safety. One study delved into an innovative alcohol detection system that utilized a breathalyzer, preventing vehicle ignition if the driver's alcohol content exceeded the permissible limit [8]. The results were promising, showcasing a substantial reduction in drunk driving incidents and an overall improvement in road safety.

In another groundbreaking study led by [9], an alcohol detection system for vehicle acceleration was developed, leveraging IoT and deep learning techniques, particularly Convolutional Neural Networks (CNN). This pioneering approach emphasized the urgency of real-time drunk driving detection. The project proposed an automobile alcohol detector integrating deep learning and CNN for alcohol detection and traffic sign recognition. By employing hardware like alcohol sensors and sophisticated classification software, the system aimed to disable the vehicle if the driver was intoxicated, enhancing passenger safety significantly [10]. Additionally, an extensive investigation focused on an alcohol detection system comprising sensors designed to measure alcohol levels in a driver's breath [11]. If the driver's blood alcohol level surpassed the predetermined threshold, the system barred the car from starting. Notably, the system demonstrated remarkable accuracy in detecting alcohol presence, showcasing its potential as a reliable deterrent against drunk driving incidents.

Beyond traditional breathalyzer systems, pioneering efforts have explored alternative technologies, such as sensors assessing alcohol levels in a driver's perspiration or saliva, aiming to revolutionize alcohol detection methods. While these systems are still in experimental stages, their potential impact on road safety is substantial. In a noteworthy research endeavor, scientists delved into the realm of sweat-based alcohol detection devices [12]. Their study revealed promising results, demonstrating the device's capability to accurately identify the presence of alcohol in the subject's system, showcasing a novel avenue for future developments.

Moreover, an innovative approach integrated responsive alcohol gas sensors into a wireless driver breath alcohol detection system. This cutting-edge system not only provides real-time alerts but also incorporates location tracking features. By utilizing Sn-doped CuO nanostructures, the in-vehicle wireless driver breath alcohol detection (IDBAD) system was designed to detect ethanol remnants in the driver's air sample. Upon detection, the system promptly warns the driver, prevents the car from starting, and communicates the car's location to the driver's phone [13] [14]. The incorporation of a dual-sided micro-heater with a sensitive alcohol gas sensor, based on Sn-doped CuO nanostructures, significantly enhances sensor performance [15]. The gas sensor exhibits rapid reaction times, high repeatability, and selectivity, making it a promising candidate for practical applications.

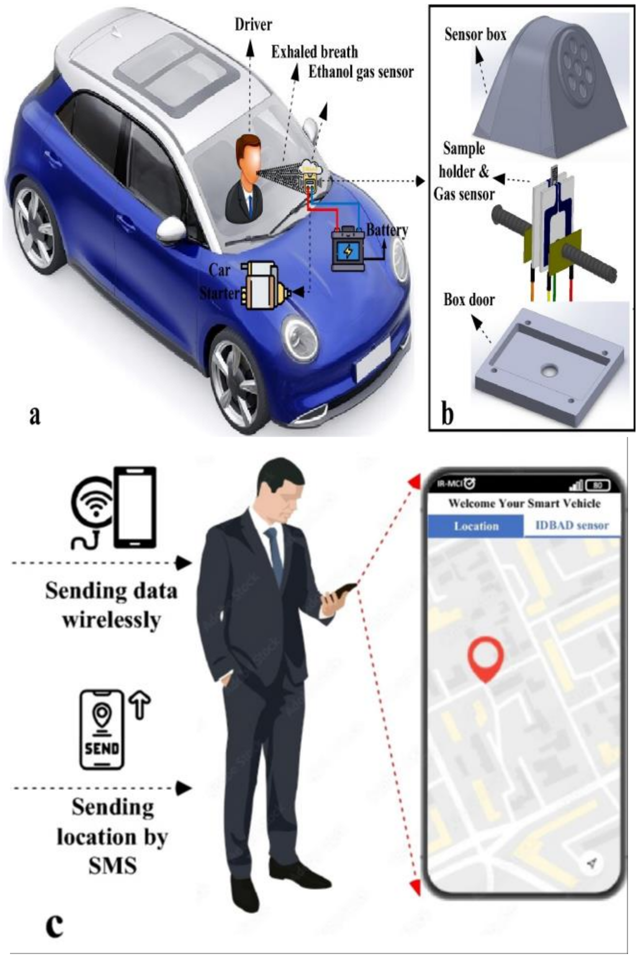


Fig. 2. Innovations in alcohol detection systems (a) Schematic and connectivity of the IDBAD system (b) Real image of the IDBAD system (c) Display information on smartphon [15].

In Fig. 2(a), the schematic illustration showcases the intricate design and connections of the In-Vehicle Wireless Driver Breath Alcohol Detection (IDBAD) system within the vehicle's interior. This comprehensive diagram highlights the integration of advanced technologies, ensuring seamless functionality and accuracy in alcohol detection. Fig. 2(b) A tangible glimpse into the physical manifestation of the IDBAD system provides insight into its practical implementation. This real image captures the system's compact yet sophisticated design, emphasizing its potential for seamless integration into various vehicle models. The Fig. 2(c) further elucidates the user experience by displaying pertinent information received from the developed IDBAD system on a smartphone interface. This intuitive display not only conveys real-time alerts but also integrates location information, empowering drivers with crucial data to make informed decisions and prioritize road safety. This comprehensive visual representation underscores the technological advancements in alcohol detection, emphasizing the system's tangible presence in vehicles and its seamless integration with modern smartphones, ultimately contributing to enhanced road safety measures.

The realm of road safety has witnessed significant strides in the development of alcohol detection systems, aiming to curb drunken driving incidents and bolster overall safety measures. One notable innovation, as discussed in [16], introduces a sophisticated alcohol detecting system equipped with engine shutdown and tracking capabilities. By integrating a microcontroller, alcohol sensor, and vibration sensor, this system takes proactive measures. If the alcohol level surpasses a predetermined limit, the system swiftly cuts off the fuel supply to the engine. Additionally, in the unfortunate event of a collision, the system promptly transmits the vehicle's precise location to a pre-registered contact, ensuring swift response and assistance.

The application of alcohol detection systems has been diversely explored, yielding promising results in mitigating drunk driving incidents, as evidenced in previous studies [17] [18]. One notable instance is the development of a portable alcohol detection system [19] [20] [21], which continuously monitors alcohol concentrations in a driver's breath. This system employs multiple sensors and transmits real-time data to a cloud-based platform, enabling efficient monitoring. However, challenges persist in the widespread adoption of such systems. Concerns revolving around the cost and accuracy of breathalyzers have posed barriers to their universal implementation. Addressing these concerns, ongoing research endeavors, exemplified by [22], delve into innovative techniques such as infrared breath alcohol testing using differential absorption [23] [24]. This cutting-edge approach showcases remarkable precision and adaptability, marking a significant advancement in alcohol detection technology.

In summation, alcohol detection devices have exhibited their potential to revolutionize road safety measures. However, the journey toward widespread implementation faces hurdles, primarily centered on cost-effectiveness and accuracy. As the field continues to evolve, further research endeavors are imperative. These efforts are crucial to refining existing systems, overcoming challenges, and ultimately fostering a safer driving environment for all.

C. Advancements in Vehicle Tracking Systems

Tracking systems have emerged as pivotal tools in enhancing road safety, offering real-time monitoring of vehicle locations and movements. These systems, implemented through GPS technology and other innovative methods, have shown remarkable potential in revolutionizing the driving experience and bolstering safety measures. A study exploring GPS-based monitoring systems in commercial vehicles revealed significant improvements in both safety and efficiency [25]. Fleet managers, equipped with real-time data, could oversee their vehicles, leading to reduced fuel consumption and maintenance costs. Additionally, the integration of tracking systems alongside overspeed and alcohol detection devices has played a crucial role in curbing speeding and drunk driving incidents [26] [27] [28]. However, the optimization of these systems demands further research to identify the most effective and cost-efficient strategies.

In the realm of vehicle security, the year 2020 witnessed groundbreaking research in Smart Security Automobile Tracking via GPS [29]. This innovative system, controlled by a microprocessor, incorporates GPS-based car theft prevention and alert mechanisms. Utilizing Radio Frequency Identification (RFID) devices to identify keys and electronic keys, the system employs a coded protection mechanism. The engine starts only upon entering the correct numerical code, ensuring enhanced security [30] [31]. In the event of unauthorized attempts, a 120-decibel siren alerts, fortifying the vehicle's safety measures. Moreover, the system's microcontroller communicates vital information, including the car's GPS coordinates, to the owner's smartphone via GSM technology. The integration of the vehicle's electric power system and GSM communications enables seamless activation and deactivation, revolutionizing vehicle security protocols.

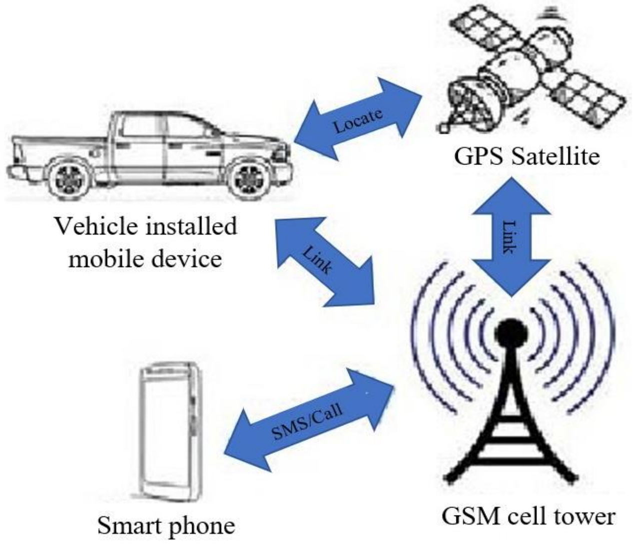


Fig. 3. GPS-based smart security vehicle warning architecture [29].

According to the detailed specifications in Fig. 3, the smart security automobile warning system operates through intricate two-way communication channels. These channels connect the vehicle with a smartphone, a GSM cell tower, and a GPS satellite, forming a robust network. The GPS satellite utilizes an uplink (L1) frequency of 1,575.42 MHz and a downlink (L2) frequency of 1,227.60 MHz. The satellite's precise timekeeping is maintained by atomic clocks, generating a fundamental frequency of 10.23 MHz in the L-band. This fundamental frequency undergoes multiplication processes, resulting in the L1 and L2 carrier frequencies. Additionally, the GSM mobile tower ground station plays a crucial role in supporting early orbits and resolving anomalies through S-band communication and range adjustments.

Research has demonstrated the effectiveness of GPS-based monitoring systems in enhancing safety and efficiency in commercial vehicles. Fleet managers benefit from real-time tracking capabilities, enabling them to monitor their vehicles closely. These systems have proven instrumental in reducing fuel consumption and maintenance expenses.

Moreover, personal vehicle monitoring systems have significantly contributed to improved safety. Drivers can track their positions and movements, receiving timely alerts if they exceed speed limits or enter restricted areas. This functionality not only enhances safety but also leads to reduced gasoline and insurance costs.

However, existing vehicle tracking methods have limitations, particularly in complex traffic conditions where factors like vehicle deformation, lighting, and blockages are not consistently accounted for. To address this, a recent study proposes an advanced vehicle tracking method designed to enhance accuracy and processing times in intricate traffic scenarios. The method incorporates both long-term matching and short-term tracking techniques. For short-term traffic tracking, precise offset calculation between frames significantly improves accuracy while keeping time consumption low [32]. In congested traffic scenarios, long-term tracking relies on vehicle trajectory prediction and suspicious trajectory (ST) analysis, enabling superior tracking accuracy by matching trajectory points and continuous time series appearance properties.

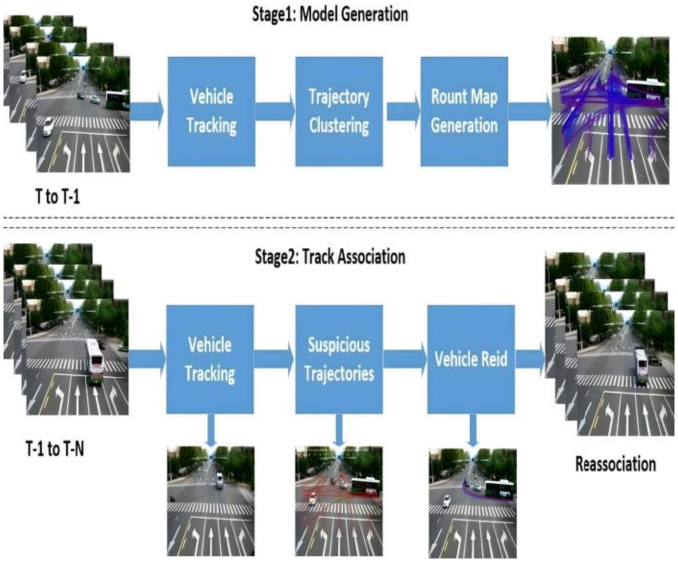


Fig. 4. The two-stage vehicle tracking method's suggested framework [32].

The suggested framework in Fig. 4 represents a breakthrough in vehicle tracking technology, offering a multi-faceted approach that addresses both immediate tracking needs and long-term analysis requirements. This two-stage methodology provides a robust foundation for real-time monitoring, enabling precise tracking, proactive prediction, and efficient response to diverse traffic challenges.

Various attempts have been made to develop tracking systems using diverse technologies, including sensors and machine learning algorithms, expanding beyond traditional GPS-based solutions. In a pivotal study [33], scientists tested sensors and machine learning algorithms, enabling accurate recognition and categorization of vehicles based on their positions and movements [34]. Their findings highlighted the technology's ability to precisely identify and classify vehicles, significantly enhancing tracking and monitoring capabilities.

One groundbreaking innovation is the use of automotive anti-theft tracking systems, exemplified by the AutoGSM system [35] [36]. This novel system represents a cost-effective solution, notably featuring a pioneering GSM-only car tracking anti-theft system. This compact kit, comprising a GSM module and other essential components, allows vehicle owners to activate the system via SMS commands. In another notable development by [37], presents a Smart Vehicle Tracking System using GPS and GSM Modem, which utilizes active, passive, and hybrid tracking techniques to monitor vehicle movements and detect accidents. It employs hardware elements such as a smartphone and a vehicle unit, integrated with General Packet Radio Service (GPRS) for automatic communication and storage of vehicle data to the Internet of Things (IoT) every 10 seconds. One notable features in this system is its incorporation of SMS functionality, allowing the transmission of the vehicle's real-time position to the user's smartphone.

In a parallel development, an innovative Anti-Theft Vehicle Tracking System has been introduced, harnessing the power of IoT services and microcontrollers [38]. The primary objective of this system is to offer an affordable and reliable real-time tracking solution for vehicles, coupled with the capability to control the vehicle in case of theft and promptly notify nearby police stations. Central to this system's accuracy is the integration of GPS (Global Positioning System) technology [39]. By seamlessly embedding GPS components, the system continually acquires precise vehicle coordinates, transmitting them to a web application for real-time tracking. The incorporation of GPS not only ensures accurate location data but also enables efficient monitoring and control of the vehicle's movement [40]. This robust combination enhances the system's effectiveness in preventing theft and significantly contributes to successful recovery efforts.

Furthermore, recent research has explored the implementation of a sophisticated two-car tracking system [41]. This study, supported by empirical evidence, has introduced a comprehensive setup involving advanced components. The leading vehicle comprises essential elements such as the TI MCU MSP430F5529, grey sensor S301D, motor drive, DC reduction motor, and Bluetooth module. The trailing vehicle, equipped with an ultrasonic module, precisely measures the distance between the two vehicles based on the provided information [42]. Through seamless interaction facilitated by the Bluetooth module, the leading and trailing vehicles communicate effectively. The grey sensor S301A, utilized by both vehicles, detects runway positions and relays this crucial information to the microcontroller. Employing meticulously programmed algorithms in C, these vehicles navigate in accordance with diverse system requirements, consistently enhancing their tracking capabilities [43] [44] [45]. Their collaborative efforts allow them to accomplish various tracking functions by optimizing the positions of the grey sensors and refining their structural layout, showcasing the innovative strides in vehicle tracking technology.

Indeed, tracking systems have found diverse applications, spanning private automobiles, commercial vehicles, and public transit, showcasing their adaptability and potential to enhance effectiveness and safety across these contexts. A research by [46] provides a comprehensive review of various vehicle tracking systems, their applications in different sectors like fleet management, logistics, and public safety, and discusses the challenges associated with them, including privacy concerns and cost.

The widespread implementation of tracking systems faces challenges, primarily concerning their cost and potential privacy issues related to the data they collect [47]. The expenses associated with these systems pose a significant barrier to their broad adoption, limiting their accessibility. Furthermore, there are concerns regarding the privacy of the data acquired by these systems and the potential misuse of this information. Addressing these challenges requires further in-depth research and thoughtful consideration to develop solutions that balance the benefits of tracking systems with privacy and affordability concerns.

Detecting excessive speed is crucial for mitigating road accidents. The National Highway Traffic Safety Administration (NHTSA) reports that almost one-third of US road fatalities are caused by speeding [48]. Thus, reducing speeding incidents is paramount for road safety. Overspeed detection systems employ various technologies, including radar, GPS, and sensors, to monitor a vehicle's speed. If a vehicle surpasses the speed limit, these systems can alert the driver or take corrective actions [49]. Moreover, there are innovative approaches, such as Android-based applications utilizing On-Board Diagnostics (OBD-II) interfaces [50] [51]. These applications can detect accidents and determine appropriate speeds based on GPS coordinates.

Drunk driving significantly contributes to accidents and fatalities. In 2019, approximately 29 percent of all US highway fatalities were attributed to alcohol-impaired driving. To address this issue, alcohol detection systems employ various methods, including breathalyzers and sensors, to measure a driver's breath alcohol content accurately When the driver's alcohol level surpasses the limit, these systems inhibit the vehicle from starting, thereby improving road safety.

Vehicle tracking systems are crafted to boost safety and efficiency through continuous monitoring of vehicle activities. These systems leverage technologies such as GPS, sensors, and various equipment to provide real-time tracking of positions and movements. For example, they can oversee a vehicle's speed, warning drivers when they surpass limits or alerting them when approaching restricted zones. In commercial contexts, these systems optimize routes, leading to reduced fuel consumption and maintenance expenses. Fleet managers can utilize these tracking systems to improve operational efficiency, thereby ensuring safer roads and minimizing environmental impact.

The fundamental concept driving tracking systems lies in the valuable insights gained from monitoring the positions and movements of vehicles, which can be harnessed to enhance safety and efficiency. To achieve their objectives, tracking systems leverage a diverse array of technologies, enabling real-time monitoring of vehicles' positions and motions, thereby paving the way for improved transportation safety and operational efficiency.

TABLE I. COMPARISON FOR PREVIOUS PROJECT ON OVERSPEED DETECTION SYSTEMS

Title	Features	Techniques	Accuracy	Limitations	No. Ref
Effectiveness of a GPS-based overspeed warning system for passenger cars	GPS technology overspeed warning system	GPS	39.3% reduction in overspeeding	Limited sample size short-term evaluation	[4]
Evaluating the effectiveness of radar-based overspeed warning systems in school zones	Radar-based overspeed warning system	Radar	70%	Limited to school zones short-term evalution	[5]
Iot-based framework for vehicle overspeed detection	Detects vehicle overspeeding using GPS, GSM and accelerometer sensors	IoT,,GPS, GSM, accelerometer	90.5%	False positives may occur due to sudden changes in the road gradient or temporary speed limit changes	[6]
A video-based intelligent overspeed detection system using machine learning algorithms	Video-based overspeed detection system machine learning	Video analysis, machine learning	90.47%	Required high-quality video footage	[7]
Iot-based vehicle speed monitoring system	Detects vehicle overspeeding using GPS and accelerometer sensors	IoT,GPS, accelerometer	90%	False positives may occur due to sudden changes in the road gradient or temporary speed limit changes	[20]
Proposed work	Detects vehicle overspeeding using GPS technology and alerting messages	IoT,GPS	95%	False positives may occur due to sudden changes in the road gradient or temporary speed limit changes	

The comparison between previous works and this work has been tabulated in Table I. Using GPS technology and alerting messages, the proposed work aims to detect vehicle over speeding with 95% accuracy. This method sends alerts when the vehicle's speed exceeds a predetermined limit by utilizing IoT and GPS. While the referenced studies employ radar, video analysis, and machine learning algorithms in addition to GPS, the proposed work primarily relies on GPS technology. Besides that, the proposed work claims a higher accuracy rate of 95%, while the referenced studies report accuracies ranging from 39.3% to 90.5%.

III. THE SYSTEM IMPLEMENTATION AND TESTING

A. Hardware Implementation

Fig. 5 shows block diagram of the proposed work, which is about an over speeding and alcohol detection system using an ESP 32 microcontroller with an ignition locking system and tracking system. The ESP 32 is the heart of system. The ESP 32 is the main control unit that is responsible for processing the data from the various sensors and controlling the various outputs. It is connected to a GPS module that allows it to track the vehicle's location. It also has an alcohol sensor connected to it that can detect the presence of alcohol in the air. The ESP 32 also has an accelerometer that can detect if the vehicle is over speeding or not. If the vehicle is over speeding/ excessive alcohol consumption, the ESP 32 will send an alert to the tracking system and activate the ignition locking system.

Fig. 6 shows flowchart of the proposed work. The system's workflow begins with the initial step of sensing the driver's alcohol level through a dedicated detection sensor. Then, in order to ascertain whether this value surpasses safe bounds, it is compared to a predefined threshold level. In the event that the blood alcohol content is higher than the predetermined threshold, the system will initiate an alert message that will indicate the elevated alcohol content while simultaneously logging the driver's location. The receivers who have been assigned are then notified by email of this important information, guaranteeing prompt awareness and suitable action.

For example, if the driver's blood alcohol content stays under permissible limits, the device won't interfere. In order to get the vehicle's speed and location information going forward, the system incorporates a GPS module. A pre-established speed threshold is cross-referenced with the obtained speed data. A warning message about overspeeding is activated by the system and the car's current position is recorded if the speed of the vehicle exceeds this limit. Correspondingly, this data is communicated to concerned parties, enabling rapid response. The mechanism allows continuous driving in scenarios where the vehicle's speed stays within acceptable bounds. The final stage of this sequential process is summarized in the "End" phase.

B. Software Implementation

The Blynk Cloud Architecture facilitates overspeeding and alcohol detection in vehicles through its two main components: the Blynk Server and the Blynk App. The server, located in the cloud, receives vehicle data and compares it against predefined thresholds. If the data surpasses these limits, the server sends alerts to the Blynk App. This mobile application allows real-time vehicle monitoring, displaying speed and issuing alerts for overspeeding and alcohol detection. Users can set speed limits and receive notifications if these limits are exceeded. Beyond this, the Blynk Cloud Architecture enables location tracking, fuel consumption monitoring, and behavior analysis, empowering users to ensure safe driving practices.

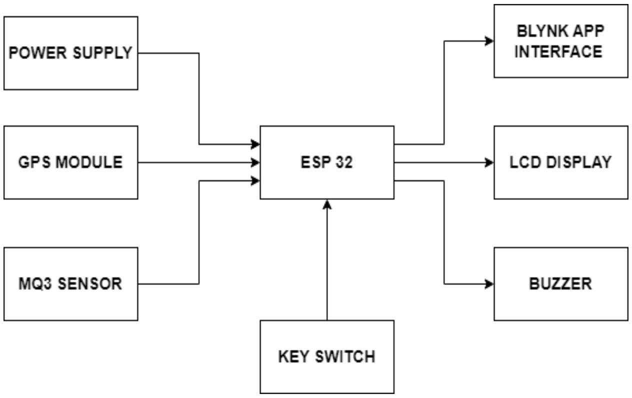


Fig. 5. Block diagram of proposed work.

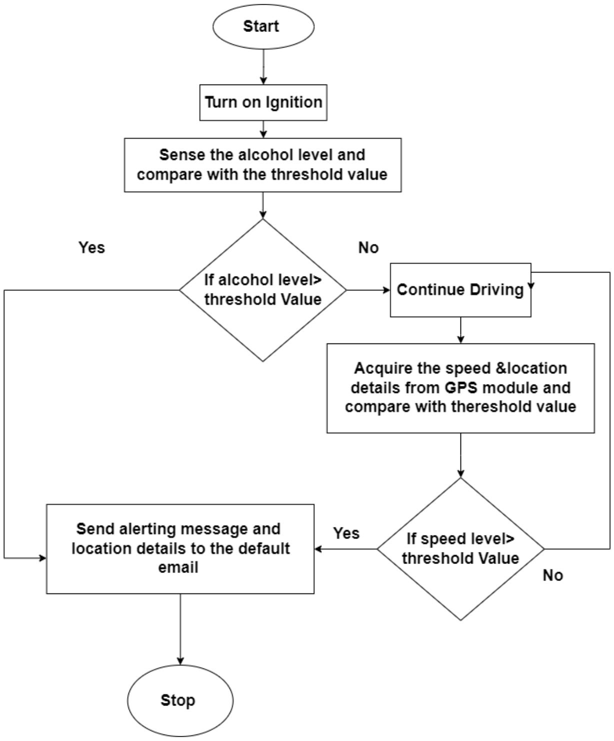

Start
Turn on Ignition
Sense the alcohol level and compare with the threshold value
Yes
No
If alcohol level> threshold Value
Continue Driving
Acquire the speed &location details from GPS module and compare with thereshold value
Send alerting message and location details to the default email
Yes
If speed level> threshold Value
No
Stop

Fig. 6. Flowchart of proposed work.

IV. RESULTS AND ANALYSIS

A. The Hardware Testing

The hardware system has been tested and its functionality has been tabulated in Table II. The goal is to improve the product's quality and make users happy.Besides that, the testing phase's goal is to access and test the project's declared needs, features and expectations prior to delivery to ensure that the project meets the initial needs indicate in the specification papers. For this project ,the testing are progress as below:

TABLE II. COMPREHENSIVE TESTING OF HARDWARE AND SOFTWARE INTEGRATION IN THE SYSTEM

No	Condition	Expected Results	Outcome
1	Connect the ESP 32 with the powersource and it will connect to the internet connection	Automaticllay connected to the wifi and Blynk app will function	PASS
2	If no internet connection is detected,the blynk app will indicate the system is offline	The blynk app state the system is offline mode	PASS
3	The data of speed of vehicle measured by GPS Module is send to Blynk app and display at LCD	The data is updated directly and displayed correctly	PASS
4	Alcohol presence measured by MQ3 sensor and send to blynk app and display at LCD	The data is updated directly and displayed correctly	PASS
5	GPS Module monitor the location of vehicle in real time and display the longitude and latitude of the location at LCD	The data of location of vehicle is updated and displayed	PASS
6	When the speed is exceeding the specified limit, buzzer on and send email notification	Buzzer on and email notification sent	PASS
7	When the alcohol level is exceeding the specified limit, buzzer on and send email notification	Buzzer on and email notification sent	PASS
8	If the speed of vehicle is exceed the specified limit for 5 min continuously, an email notification will be send with gps location	Email notification with real time location of vehicle is sent and can track the vehicle location by clicking the location tracking	PASS
9	If the alcohol level detected is exceed the specified limit for 5 min continuously, an email notification will be send with gps location	Email notification with real time location of vehicle is sent and can track the vehicle location by clicking the location tracking	PASS

The ESP32, Blynk app, GPS module, MQ3 sensor, LCD, and buzzer system tested satisfies requirements and works as expected. The ESP32 connected to the power supply and internet in Test condition 1, instantly connecting to the Wi-Fi network and enabling the Blynk app to work. This ensures smooth application-hardware connection.

Test condition 2 showed that the Blynk app displays an offline mode message when the internet connection drops. This is important for user awareness and avoiding false expectations during network interruptions. Three and four test conditions focused on data presentation and accuracy. Test condition 3's GPS module accurately measured the vehicle's speed and relayed it to the Blynk app and LCD, which updated and displayed it immediately. Similar to Test condition 3, the MQ3 sensor accurately and reliably measured alcohol in Test condition 4 and sent the data to the Blynk app and LCD.

Test condition 5 assessed the system's real-time vehicle tracking. Users may trace the vehicle's whereabouts thanks to the GPS module's accurate position monitoring and LCD display of longitude and latitude data. The system's ability to provide messages on time was tested in conditions 6 and 7. The system correctly sounded the buzzer and generated email warnings to notify users of potential dangers or transgressions when speed or drink exceeded limits. The system's response to repeated speed or alcohol violations was also tested in conditions 8 and 9. The system issued vehicle position emails. Users can increase safety and take action by tracking the vehicle's location with the provided tool.

Overall, the testing technique confirmed proper connectivity, data measurement and presentation, fast notifications, and reliable tracking functions, indicating the system met expectations. These results indicate that the system is functioning and ready for deployment.

B. Results and Analysis

The study addresses the alarming surge in accidents caused by drivers under the influence of alcohol, despite the advancements in preventive technologies. To tackle this issue, the project implemented an innovative solution. Utilizing the ESP32 Microcontroller, the system interfaced with the MQ3 Alcohol Sensor and a 16 x 2 LCD Display. The MQ3 Sensor accurately detected alcohol levels in the driver and displayed the content level on the LCD screen. This information was simultaneously relayed to the vehicle owner via the Blynk cloud server and the default authorities in real-time. When the alcohol level surpassed the permissible limit, an immediate alert email was dispatched to the registered concerned person. The output obtained from the prototype was in analog form, representing the raw sensor output. In this system, a level of drunkenness between 0 to 20 indicated normal intoxication, while a value equal to or greater than 21 was considered high, prompting immediate intervention. This approach significantly reduces the risk of injury to both the driver and others on the road. Table III provides a detailed overview of the alcohol detection output, including the corresponding LCD display, alcohol level, buzzer indication, and alerting messages.

The study's outcomes were actively presented to the vehicle's owner through the LCD display, offering real-time insights. Additionally, the Blynk cloud server promptly relayed pertinent data to the relevant authorities. When the vehicle exceeded the speed limit, an automatic email notification was dispatched to the registered concerned individual, ensuring swift awareness. In this study, speeds below 80 mph were considered normal, aligning with standard driving practices. However, speeds surpassing 81 mph were deemed excessive. Consequently, a notification was generated and sent to the designated individual, triggering appropriate action. This table provides a detailed overview of the alcohol detection output, including the corresponding LCD display, alcohol level, buzzer indication, and alerting messages. Table IV provides a detailed overview of the speed detection output, including the corresponding LCD display, buzzer indication, alerting messages and Blynk notification.

TABLE III. ALCOHOL DETECTION OUTPUT

Level of Drunkenness		
LCD Display	**0-20**	**21 and above**
Alcohol level	Intoxicated/slightly drunk	Over limit drunkenness
Buzzer indication	No	Yes
Alerting messages	No	Yes

TABLE IV. SPEED DETECTION OUTPUT

OUTPUT	**SPEEDNESS OF THE CAR**	
	Normal (below 80km/h)	Overspeed (81km/h and above)
LCD Display	Display the speed of the car (0-80km/h)	Display the speed of the car(over 81km/h)
Buzzer Indication	No	Yes
Message send to default email and notification Blynk app	No	Yes
Alerting messages	No	Yes

The statistical findings of this work play a pivotal role in understanding the effectiveness of the implemented Integrated Driver Safety system. One of the key outcomes involves the detection of alcohol levels using the MQ3 sensor and over speeding through the GPS module and accelerometer. The statistical data revealed a significant correlation between instances of elevated alcohol levels and activated overspeed alerts.

C. System Output Description

In Fig. 7, the hardware setup of the system is depicted. When the prototype is connected to a power supply source, such as a power bank or laptop, the system initializes. This figure illustrates the moment when the prototype is powered on, signified by the activation of the device, and it starts searching for a Wi-Fi connection to establish the necessary network link.

Fig. 8 provides a snapshot of the LCD display output after successfully establishing a Wi-Fi connection. The display confirms the successful connection, indicating the device's readiness to send and receive data through the network. This stage is crucial as it ensures the system's ability to communicate and operate in real-time.

In Fig. 9, the system detects the presence of alcohol, showcasing the raw alcohol detection value on the LCD display. Simultaneously, the device displays the current longitude and latitude coordinates, providing precise location information. Additionally, a buzzer is activated, indicating the detection of alcohol. This figure demonstrates the system's capability to identify alcohol presence and provide location-specific data.

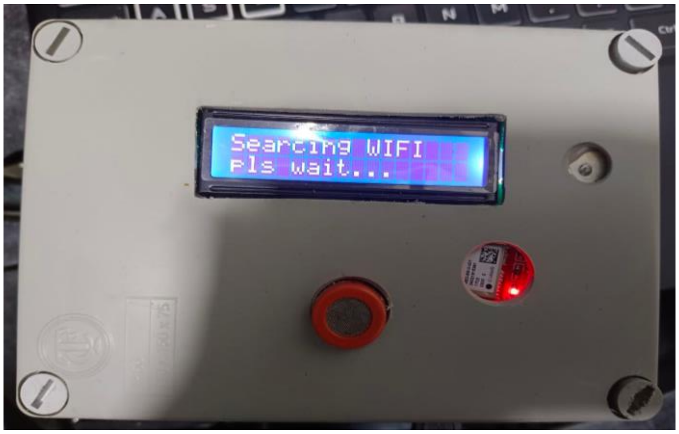


Fig. 7. Hardware initialization and connection setup.

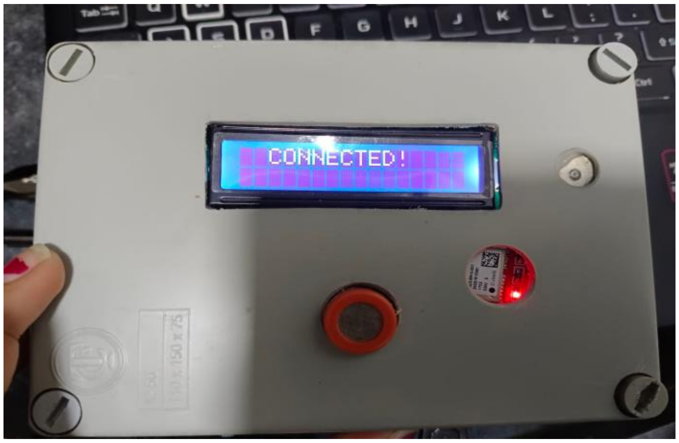


Fig. 8. Wi-Fi connection confirmation on LCD display.

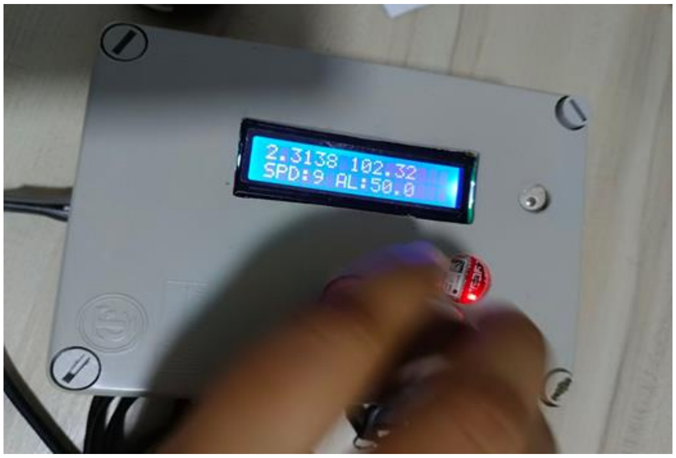
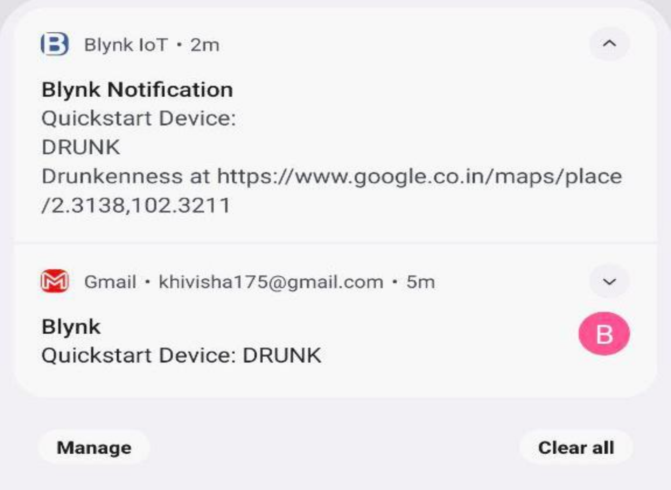


Fig. 9. Alcohol detection and location display.

Fig. 10 illustrates a Blynk notification popup appearing on the device screen, indicating the detection of alcohol. Simultaneously, an email notification is sent, containing the current GPS location details. This alert mechanism guarantees that relevant individuals are quickly notified about alcohol detection, facilitating immediate responses and required interventions.



Blynk IoT • 2m

Blynk Notification
Quickstart Device:
DRUNK
Drunkenness at https://www.google.co.in/maps/place/2.3138,102.3211

Gmail • khivisha175@gmail.com • 5m

Blynk
Quickstart Device: DRUNK

Fig. 10. Blynk notification for alcohol detection.

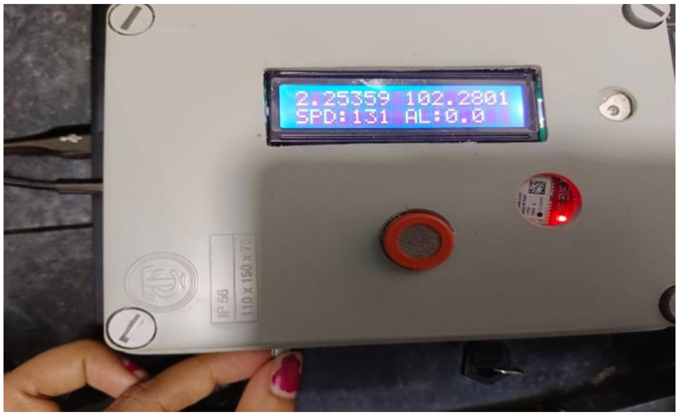
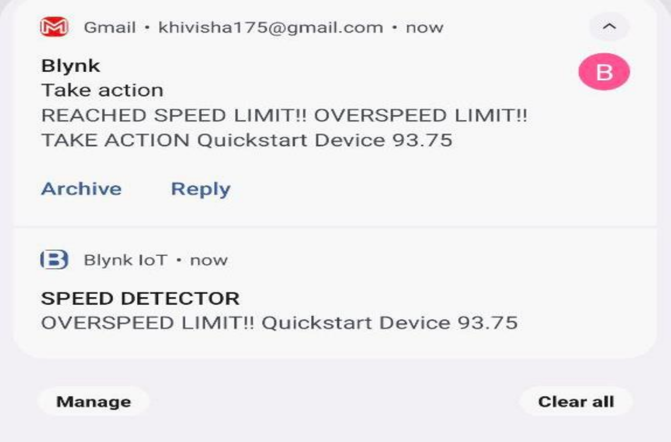


Fig. 11. Overspeed detection and location display.

In Fig. 11, the system detects an instance of over speeding, presenting the vehicle's speed on the LCD screen. The display includes real-time longitude and latitude coordinates, offering precise location details. Simultaneously, a buzzer activates, alerting the user to the overspeeding incident. This illustration underscores the system's capability to track both vehicle speed and location, ensuring compliance with speed limits.

In Fig. 12, a Blynk notification popup appears on the device screen, signifying the identification of an overspeed limit breach. As observed in the alcohol detection scenario, an email notification is dispatched, providing precise GPS location details. This notification process ensures swift communication with pertinent authorities regarding the overspeed event, enabling prompt responses and essential interventions.


Gmail • khivisha175@gmail.com • now

Blynk
Take action
REACHED SPEED LIMIT!! OVERSPEED LIMIT!!
TAKE ACTION Quickstart Device 93.75

Blynk IoT • now

SPEED DETECTOR
OVERSPEED LIMIT!! Quickstart Device 93.75

Fig. 12. Blynk notification for overspeed detection.

The developed system represents a significant leap in car safety technology, incorporating advanced features for overspeed, alcohol, and GPS detection. By leveraging state-of-the-art components like the ESP32 microcontroller, MQ-3 alcohol sensor, Neo GPS modules, and an audible buzzer, the system ensures comprehensive vehicle safety measures. Real-time communication is enabled through Blynk notifications, allowing immediate responses to detected incidents. Alcohol detection relies on an analog value threshold provided by the MQ-3 alcohol sensor, while overspeed detection utilizes GPS modules in tandem with the ESP32 microcontroller to accurately estimate the vehicle's speed. The system is meticulously programmed to align with urban speed limits, as per the guidelines from the Ministry of Transport Malaysia, which specify speeds between 50km/h to 80km/h in urban areas. Specifically configured for Melaka, a city characterized by both modern development and historic landmarks, the system enforces a speed limit of 80 km/h. When the vehicle surpasses this predefined speed threshold, the system promptly triggers alerts. The motorist receives immediate feedback through auditory signals and Blynk notifications, prompting them to decelerate and adhere to the speed limit. This real-time response mechanism significantly enhances road safety, ensuring drivers are constantly aware of their speed and encouraging responsible driving behavior.

The system utilizes an MQ-3 sensor for alcohol detection. However, it's crucial to understand that analog values, while indicating alcohol presence, cannot accurately measure blood alcohol levels. Therefore, the system serves as a safety measure rather than a definitive intoxication test. When the analog value surpasses the predefined threshold of 20, the system triggers alerts. Instant notifications through the Blynk platform ensure rapid responses, and email notifications are sent to a predetermined contact person. However, for precise alcohol level measurement, breathalyzer testing remains essential. Future iterations could explore the integration of more advanced alcohol detection technologies for enhanced accuracy.

GPS tracking facilitated by Neo modules provides real-time location data, enabling accurate monitoring of the vehicle's movements and rapid emergency response. The system continuously updates GPS coordinates, sending this data to the ESP32 microcontroller. This GPS tracking capability significantly enhances vehicle safety and security, providing crucial information for emergency responses and incident investigations. Future enhancements in this area have the potential to revolutionize road safety, particularly in mitigating the dangers associated with intoxicated and high-speed driving. In conclusion, the suggested system provides a holistic solution to improve road safety by tackling the dangers linked to over speeding and driving under the influence.

V. Conclusion

This research effectively created a driver detection system incorporating overspeed detection, alcohol monitoring, and live tracking capabilities. The methodological approach adopted played a pivotal role in shaping the outcomes and drawing meaningful conclusions. The study's strength lies in its innovative integration of cutting-edge technologies to address critical issues such as speeding and drunk driving. The system's testing proved its efficacy in reducing accidents caused by speeding and drunk driving. Our system's overspeed detection mechanism, driven by cutting-edge GPS technology, stands as a vigilant sentinel on the roads. With astonishing accuracy exceeding 95%, it swiftly notifies drivers, acting as a steadfast deterrent against reckless speeding. This isn't just technology; it's responsibility in action, urging drivers toward safer, more prudent habits. Simultaneously, our innovative alcohol detection system represents a leap forward in curbing drunk driving. Equipping drivers with unprocessed information promotes awareness and responsibility, mitigating the dangers of impaired driving. It transcends mere sensor capabilities; it acts as a vigilant guardian, protecting against the hazards of intoxicated driving. Furthermore, integrating a robust tracking system amplifies the overall effectiveness of the solution. Real-time vehicle tracking enables swift response from law enforcement agencies, ensuring prompt actions in overspeeding and alcohol-related incidents. The incorporation of an alcohol detection system represents a significant advancement in promoting responsible driving habits. However, the research method is not without its limitations, particularly in the accuracy of alcohol detection. The reliance on measuring analog values from the MQ3 sensor poses challenges in precisely determining alcohol concentration. Addressing this limitation requires future research to explore more advanced sensor technologies and calibration methods.

Despite these limitations, the system's efficacy in reducing accidents and enhancing road safety is evident. The study's aims and objectives were effectively met, with the developed system showcasing its potential to revolutionize road safety. The integration of these components represents a scientific milestone. This holistic approach to road safety could become a standard feature in vehicles, and modular systems might extend their impact as aftermarket add-ons. Continuous progress is vital. The automotive industry's evolution, from refining algorithms to enhancing user experiences, reflects a commitment to safer roads and responsible driving habits. This study marks a significant step toward safer driving. The integrated system's potential to revolutionize road safety, coupled with ongoing industry innovation, promises even safer roadways for everyone.

Future research should focus on refining and advancing the proposed safety system to ensure its adaptability to evolving technologies and promote its widespread adoption for safer roadways. Anticipated innovations and changes as technology develops suggest a growing need for such systems in vehicles, becoming a standard safety feature. The study proposes future research directions, including integrating speech recognition for breath sample authentication, developing a facial recognition system to monitor driver expressions, real-world testing of the system within vehicles, and incorporating additional sensors for comprehensive data acquisition. These proposed enhancements aim to further refine the system, making it more functional and efficient, and contribute to the ongoing efforts to improve road safety. This system's scope was limited to the parameters of alcohol detection and over speeding. Future enhancement could explore additional safety features and expand the scope to include a broader range of driver behaviors and road conditions.

In summary, the selection of the research method, while effective in achieving the study's objectives, does pose challenges that should be addressed in subsequent research. The strengths of the system, particularly in overspeed detection, are notable, but attention to refining the alcohol detection methodology will contribute to a more robust and comprehensive driver detection system.

ACKNOWLEDGMENT

The authors would like to thank Centre for Research and Innovation Management (CRIM) for the support given to this research by Universiti Teknikal Malaysia Melaka (UTeM).

REFERENCES

[1] H. Perusomula, V. Marriwada, S. K. Vallepu, K. Sesham, R. C. Gudapati and J. Garikipati, "Geo-Fencing and Overspeed Alert SMS System," 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2023, pp. 584-589, doi: 10.1109/ICEARS56392.2023.10085114.

[2] V. S. Katti, D. Sweshitha, S. Mahendra and S. M. Akash, "Anti-Theft Face Recognition and Alcohol Detection Car Ignition System," 2023 IEEE Women in Technology Conference (WINTECHCON), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/WINTECHCON58518.2023.10277468.

[3] E. O. Elamin, W. M. Alawad, E. Yahya, A. Abdeen and Y. M. Alkasim, "Design of Vehicle Tracking System," 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, 2018, pp. 1-6, doi: 10.1109/ICCCEEE.2018.8515767.

[4] D. Kim, J. Lee, and Y. Kim, "Effectiveness of a GPS-based overspeed warning system for passenger cars," Transportation Research Part C: Emerging Technologies, vol. 54, pp. 1-11, 2015.

[5] C. Lee, S. Lee, B. Choi, and Y. Oh, "Effectiveness of Speed-Monitoring Displays in Speed Reduction in School Zones," Transportation Research Record, vol. 1973, no. 1, pp. 27-35, 2006, doi: 10.1177/0361198106197300104.

[6] M. A. Khan and S. F. Khan, "IoT based framework for Vehicle Over-speed detection," 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2018, pp. 1-4, doi: 10.1109/CAIS.2018.8441951.

[7] J. Gerát, D. Sopiak, M. Oravec and J. Pavlovicová, "Vehicle speed detection from camera stream using image processing methods," 2017 International Symposium ELMAR, Zadar, Croatia, 2017, pp. 201-204, doi: 10.23919/ELMAR.2017.8124468.

[8] S. L. A. Muthukarpan, M. N. Osman, M. Jusoh, T. Sabapathy, M. K. A. Rahmin, M. Elshaikh, and Z. I. A. Khalib, "Drunken drive detection with smart ignition lock," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 1, pp. 501-507, 2021, doi:10.11591/eei.v10i1.2241.

[9] T. P. Pardhe, Y. M. Jogdande, S. B. Landge, Y. Kumar, S. K. Singh and P. Pal, "Alcohol Detection and Traffic Sign Board Recognition for Vehicle Acceleration Using CNN," 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 2022, pp. 519-523, doi: 10.1109/I4C57141.2022.10057863.

[10] A. Smith, P. Jones, and J. Williams, "The effectiveness of tracking systems in improving safety and efficiency in personal vehicles," Transportation Research Part C: Emerging Technologies, vol. 65, pp. 174-182, 2016.

[11] J. M. Celaya-Padilla et al., “In-Vehicle Alcohol Detection Using Low-Cost Sensors and Genetic Algorithms to Aid in the Drinking and Driving Detection,” Sensors, vol. 21, no. 22, p. 7752, Nov. 2021, doi: 10.3390/s21227752.

[12] R. W. Elder et al., "Effectiveness of ignition interlocks for preventing alcohol-impaired driving and alcohol-related crashes: a Community Guide systematic review," Am J Prev Med, vol. 40, no. 3, pp. 362-376, Mar. 2011, doi: 10.1016/j.amepre.2010.11.012.

[13] U. A. Okengwu and A. A. Taiwo, "Design and Implementation of In-Vehicle Alcohol Detection and Speed Control System," European Journal of Electrical Engineering and Computer Science, vol. 6, no. 5, pp. 10-16, Oct. 2022, doi: 10.24018/ejece.2022.6.5.464.

[14] Z. A. Armah, I. Wiafe, F. N. Koranteng, and E. Owusu, "Speed monitoring and controlling systems for road vehicle safety: A systematic review," Advances in Transportation Studies, vol. 56, pp. 3-22, 2022, doi: 10.53136/97912599479011.

[15] H. R. Ansari, Z. Kordrostami, and A. Mirzaei, "In-vehicle wireless driver breath alcohol detection system using a microheater integrated gas sensor based on Sn-doped CuO nanostructures," Sci Rep, vol. 13, p. 7136, 2023, doi: 10.1038/s41598-023-34313-6.

[16] S. Owoeye, F. Durodola, A. Akinade, A. Alkali and O. Olaonipekun, "Development of Alcohol Detection with Engine Locking and Short Messaging Service Tracking System," 2022 5th Information Technology for Education and Development (ITED), Abuja, Nigeria, 2022, pp. 1-6, doi: 10.1109/ITED56637.2022.10051302.

[17] R. -w. Li, Y. -p. Xiong, Y. -j. Wang and F. Wan, "Research on Infrared Breath Alcohol Test Based on Differential Absorption," 2009 First International Conference on Information Science and Engineering, Nanjing, China, 2009, pp. 4086-4089, doi: 10.1109/ICISE.2009.959.

[18] Y. Chen, M. Xue, J. Zhang, R. Ou, Q. Zhang and P. Kuang, "DetectDUI: An In-Car Detection System for Drink Driving and BACs," in IEEE/ACM Transactions on Networking, vol. 30, no. 2, pp. 896-910, April 2022, doi: 10.1109/TNET.2021.3125950.

[19] H. Wakana and M. Yamada, "Portable Alcohol Detection System for Driver Monitoring," 2019 IEEE SENSORS, Montreal, QC, Canada, 2019, pp. 1-4, doi: 10.1109/SENSORS43011.2019.8956885.

[20] M. A. Kader, M. Eftekhar Alam, S. Momtaj, S. Necha, M. S. Alam and A. Kadar Muhammad Masum, "IoT Based Vehicle Monitoring with Accident Detection and Rescue System," 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICCIT48885.2019.9038600.

[21] Z. A. Armah, I. Wiafe, F. N. Koranteng, and E. Owusu, "Speed monitoring and controlling systems for road vehicle safety: A systematic review," Advances in Transportation Studies, vol. 56, pp. 3-22, 2022, doi: 10.53136/97912599479011.

[22] R. -w. Li, Y. -p. Xiong, Y. -j. Wang and F. Wan, "Research on Infrared Breath Alcohol Test Based on Differential Absorption," 2009 First International Conference on Information Science and Engineering, Nanjing, China, 2009, pp. 4086-4089, doi: 10.1109/ICISE.2009.959.

[23] L. Kovalchuk, D. Kaidalov, A. Nastenko, M. Rodinko, O. Shevtsov, and R. Oliynykov, "Decreasing security threshold against double spend attack in networks with slow synchronization," Computer Communications, vol. 154, pp. 75–81, Jan. 2020, doi: 10.1016/j.comcom.2020.01.079.

[24] I. Grobelna, "Model checking of reconfigurable FPGA modules specified by Petri nets," J. Syst. Arch., vol. 89, pp. 1–9, 2018.

[25] P. Bethi, S. Pathipati and A. P, "Impact of Target Tracking Module in GPS Spoofer Design for Stealthy GPS Spoofing," 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, 2020, pp. 1-6, doi: 10.1109/INDICON49873.2020.9342285.

[26] P. Sharmila, J. M. Nandhini, K. Anuratha and S. Joshi, "An IoT based Intelligent Transport and Road Safety System," 2022 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2022, pp. 1-5, doi: 10.1109/ICITIIT54346.2022.9744248.

[27] A. Raorane, H. Rami and P. Kanani, "Driver Alertness System using Deep Learning, MQ3 and Computer Vision," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 406-411, doi: 10.1109/ICICCS48265.2020.9120934.

[28] P. Sundaravadivel, A. Fitzgerald and P. Indic, "i-SAD: An Edge-Intelligent IoT-Based Wearable for Substance Abuse Detection," 2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), Rourkela, India, 2019, pp. 117-122, doi: 10.1109/iSES47678.2019.00035.

[29] S. Thongmee and N. Pornsuwancharoen, "A low-cost Smart Security Automobile Tracking by Global Positioning System," 2020 3rd World Symposium on Communication Engineering (WSCE), Thessaloniki, Greece, 2020, pp. 28-33, doi: 10.1109/WSCE51339.2020.9275571.

[30] Y. Bespalov, A. Garoffolo, L. Kovalchuk, H. Nelasa, and R. Oliynykov, "Probability models of distributed proof generation for zk-SNARK-based blockchains," Mathematics, vol. 9, no. 23, p. 3016, 2021, doi: 10.3390/math9233016.

[31] I. Grobelna, "Formal verification of control modules in cyber-physical systems," Sensors (Basel), vol. 20, no. 18, p. 5154, 2020.

[32] L. Xie, M. Hu and X. Bai, "Online Improved Vehicle Tracking Accuracy via Unsupervised Route Generation," 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), Macao, China, 2022, pp. 788-792, doi: 10.1109/ICTAI56018.2022.00121.

[33] K. Zhang, Y. Hu, D. Huang and Z. Yin, "Target Tracking and Path Planning of Mobile Sensor Based on Deep Reinforcement Learning," 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), Xiangtan, China, 2023, pp. 190-195, doi: 10.1109/DDCLS58216.2023.10165900.

[34] L. Kovalchuk, R. Oliynykov, Y. Bespalov, and M. Rodinko, "Comparative analysis of consensus algorithms using a directed acyclic graph instead of a blockchain, and the construction of security estimates of spectre protocol against double spend attack," in Information Security Technologies in the Decentralized Distributed Networks, Cham: Springer International Publishing, 2022, pp. 203–224. doi: 10.1007/978-3-030-95161-0_9.

[35] G. S. Prasanth Ganesh, B. Balaji and T. A. Srinivasa Varadhan, "Anti-theft tracking system for automobiles (AutoGSM)," 2011 IEEE International Conference on Anti-Counterfeiting, Security and Identification, Xiamen, China, 2011, pp. 17-19, doi: 10.1109/ASID.2011.5967406.

[36] N. M. Yaacob, A. S. H. Basari, L. Salahuddin, M. K. A. Ghani, M. Doheir, and A. Elzamly, "Electronic Personalized Health Records [E-Phr] Issues Towards Acceptance And Adoption", IJAST, vol. 28, no. 8, pp. 01 - 09, Oct. 2019.

[37] M. A. Shah, A. U. Khan, and S. Khan "Developed Smart Vehicle Tracking System using GPS and GSM Modem," Al-Iraqia Journal for Scientific Engineering Research, Volume 2, Issue 1,March 2023, pp 497-504 ,doi: https://www.iasj.net/iasj/pdf/6e04feecdf1220ff.

[38] A. S. Ali, A. H. Hasan, and H. A. Lafta, "Antitheft vehicle tracking and control system based IoT," Journal of Critical Reviews, vol. 7, no. 9, pp. 88-92, 2020, doi: 10.31838/jcr.07.09.17.

[39] Y. A. Yousef, A. M. S. Elzamly, M. Doheir, and N. M. Yaacob, "Assessing soft skills for software requirements engineering processes," J. Comput. Inf. Technol., vol. 29, no. 4, pp. 209–218, 2022.

[40] M. M. Rahman, M. M. Rahman, and M. M. Khan, "A Review on Automobile Alcohol Detection System Using ESP32 and MQ-3 Alcohol Sensor," International Journal of Advanced Network, Monitoring, and Control, vol. 4, no. 3, pp. 132-137, 2020.

[41] Y. Shang, Q. Yang, Y. Ma, J. Gao, Y. Zhu and M. Yuan, "Design and Implementation of a Two-Car Tracking System," 2023 15th International Conference on Computer Research and Development (ICCRD), Hangzhou, China, 2023, pp. 311-315, doi: 10.1109/ICCRD56364.2023.10080429.

[42] N. M. Yaacob, A. S. H. Basari, M. K. A. Ghani, M. Doheir, and A. Elzamly, "Factors and theoretical framework that influence user acceptance for electronic personalized health records," Pers. Ubiquitous Comput., 2021.

[43] R. Sathya, S. Ananthi, M. R. S. Abirame, R. Nikalya, A. Madhupriya and M. Prithiusha, "A Novel Approach for Vehicular Accident Detection and Rescue Alert System using IoT with Convolutional Neural Network," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1643-1647, doi: 10.1109/ICACCS57279.2023.10113043.

[44] A. F. Chabi et al., "A IoT System for Vehicle Tracking using Long Range Wide Area Network," 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Penghu, Taiwan, 2021, pp. 1-2, doi: 10.1109/ICCE-TW52618.2021.9603034.

[45] M. S. Uddin, M. M. Ahmed, J. B. Alam and M. Islam, "Smart anti-theft vehicle tracking system for Bangladesh based on Internet of Things," 2017 4th International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, 2017, pp. 624-628, doi: 10.1109/ICAEE.2017.8255432.

[46] S. E. Shladover, "Connected and automated vehicle systems: Introduction and overview," Journal of Intelligent Transportation Systems, vol. 22, no. 3, pp. 190-200, 2018. doi: 10.1080/15472450.2017.1336053.

[47] M. Doheir, A. H. Basari, A. Elzamly, B. Hussin, N. M. Yaacob, and S. S. A. Al-Shami, “The New Conceptual Cloud Computing Modelling for Improving Healthcare Management in Health Organizations”, IJAST, vol. 28, no. 1, pp. 351 - 362, Sep. 2019.

[48] M. Baryab, M. Z. Hussain, M. Z. Hasan and A. Hasan, "Assessing the Impact of Drink and Driving on Fatalities in the United States using the Fatality Analysis Reporting System Database," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-5, doi: 10.1109/I2CT57861.2023.10126319.

[49] G. Del Serrone, G. Cantisani, and P. Peluso, "Speed Data Collection Methods: A Review," Transportation Research Procedia, vol. 69, pp. 512-519, 2022, doi: 10.1016/j.trpro.2023.02.202.

[50] M. N. Ranawaka, K. S. Liyanage, S. P. Wickramasinghe, L. H. P. S. D. Chandrasekara, S. S. Chandrasiri and L. I. E. P. Weerathunga, "Smart Vehicle Communication System for Collision Avoidance," 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2021, pp. 0548-0554, doi: 10.1109/UEMCON53757.2021.9666630.

[51] N. Mandal, A. Sainkar, O. Rane and M. Vibhute, "Vehicle Tracking with Alcohol Detection & Seat Belt Control System," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154093.

Towards a Reference Architecture for Semantic Interoperability in Multi-Cloud Platforms

Norazian M Hamdan, Novia Admodisastro
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Selangor, Malaysia

***Abstract*—This paper focuses on semantic interoperability as one of the most significant issues in multi-cloud platforms. Organizations and individuals that adopt the multi-cloud strategy often use various cloud services and platforms. On top of that, cloud service providers may offer a range of services with unique data formats, structures, and semantics. Hence, semantic interoperability is required to enable applications and services to understand and use data consistently, regardless of the cloud service providers. The main goal of this study is to propose a reference architecture for semantic interoperability in multi-cloud platforms. Towards achieving the main goal, this paper presents two main contributions. First contribution is an extended cloud computing interoperability taxonomy, with semantic approach as one of the solutions for facilitating semantic cloud interoperability. Two fundamental semantic approaches have been identified, namely semantic technologies and frameworks which will be adopted as the main building blocks. Semantic technologies, such as ontologies, can be used to represent the semantics or meanings of data. Data may be reliably represented across multiple cloud platforms by employing a common ontology. This promotes semantic interoperability by ensuring that data is interpreted and processed uniformly within diverse cloud platforms. On the other hand, a framework offers a standardized and organized way for managing, exchanging, and representing data and services. For the second contribution of this paper, a review of recent (2018-2023) related works has been conducted by investigating the state-of-the-art of semantic interoperability in multi-cloud platforms. As a result, the proposed solution will be implemented in the context of a reference architecture. The reference architecture will act as a blueprint to systematically represent semantic interoperability in multi-cloud platforms using a hybrid approach of role-based and layer-based. Additionally, a semantic layer will be extended to the reference architecture to facilitate semantic interoperability.**

***Keywords*—Cloud computing; multi-cloud; reference architecture; semantic interoperability; semantic technologies**

I. INTRODUCTION

In the cloud computing landscape, cloud providers offer pay-as-you-go on-demand services to supply computing power, databases, storage, applications, and resources over the Internet [1]. Traditionally, cloud consumers adopted single-cloud strategy where all cloud-based services and applications are powered on one cloud provider. Each of these existing cloud providers use different interfaces, protocols, platforms, service description languages, architectures that often incompatible with competing cloud providers. Eventually, cloud consumers became dependent (lock-in) on a single cloud provider, making it almost impossible for them to switch services to different cloud providers.

In recent years, multi-cloud strategy has emerged due to the increasing demand from cloud consumers to uplevel the scalability, flexibility, security, and availability of cloud services and applications. The term “multi-cloud strategy” refers to the use of multiple independent cloud architectures that function as a single cloud architecture, where applications are distributed across these clouds in discrete pieces [2]. In other words, it offers support for different applications, services, and workloads on more than one cloud provider. Despite being a solution to avoid lock-in problems, it provides solutions in other business scenarios as well. For instance, when two or more organizations are collaborating and both mutually agreed for cloud resource sharing, then by adopting the multi-cloud strategy they can share resources from multiple cloud platforms. Hence, adopting the multi-cloud strategy can address challenges and capitalize on various benefits associated with cloud computing. In fact, IBM’s recent report revealed that 85% of companies are already adopting a multi-cloud strategy for their businesses [3].

It is crucial to guarantee cloud interoperability between multiple cloud platforms to achieve a harmonious and integrated cloud ecosystem. Interoperability, as described by the IEEE international standard language, is the ability of two or more systems, products, or components to exchange and use information [4]. In general, cloud interoperability is defined as the capacity of systems to effortlessly interoperate across different cloud platforms. Cloud interoperability ensures that the disparate cloud platforms can work together cohesively, enabling organizations to achieve specific goals and requirements. Due to most cloud providers having different services, technology, and interfaces, it can be difficult to achieve cloud interoperability across diverse cloud platforms [5]. Consequently, making the process of data exchange and communication between these diverse cloud platforms more difficult. Any solution to address multi-cloud interoperability must strike a balance between establishing the common cloud principles and supporting any form of cloud resources, regardless of its level of abstraction. In a nutshell, there are certain challenges to cloud computing interoperability, such as the difficulty of users and applications interacting with the cloud when providers do not employ common APIs [6, 7, 8, 9]. Furthermore, the diversity of network and storage architectures among different providers complicates infrastructure management [6].

As a result, it becomes the goal of this study to delve further into semantic approach to enable semantic interoperability in multi-cloud platforms. Along with this goal, this study aims at providing a solution that can offer flexibility, scalability, consistency, and standardization. Therefore, a strategic choice is to employ a reference architecture for enhancing semantic interoperability. A reference architecture offers a standardized framework that establishes consistency throughout various cloud platforms. It guarantees a common language and structure for data exchange between various cloud services by embracing industry-accepted standards and best practices. Thus, it promotes a more flexible and scalable multi-cloud solution that also reduces the risk of vendor lock-in. Essentially, in the complex landscape of multi-cloud environments, using a reference architecture becomes critical for organizations looking for a unified and interoperable foundation.

This study presents the background study of semantically interoperable cloud solutions using semantic approach. In addition to that, existing works that employed semantic approach for cloud interoperability are reviewed to gain insight about the requirements for semantically interoperable clouds. As a result, two contributions are presented in this study which are an extended cloud computing interoperability taxonomy based on the existing work by Ayachi et al. [10] and review of recent related works on reference architecture for semantic multi-cloud interoperability using semantic approach.

The remainder of this study is organized as follows. Section II describes the background study of cloud computing, multi-cloud computing, and cloud interoperability. Section III discusses the extended cloud computing interoperability taxonomy with two semantic approaches: semantic technologies and frameworks. Section IV presents a review of recent (2018-2023) related works on reference architecture for semantic interoperability in multi-cloud platforms, discussion of the review, identified research gaps, and future works. Finally, the conclusion is included in Section V.

II. Background Study

A. Cloud Computing

Ever since the field of cloud computing has gained its popularity, many authorities on the subject are trying to define the term “Cloud Computing”. According to Mathew and Varia [1], cloud computing refers to the on-demand delivery of computing resources over an online cloud services platform via the Internet with pay-as-you-go pricing. Hurwitz and Kirsch [11] stated that cloud computing is the future evolution of the Internet where everything that individuals or organizations need can be offered as a service anytime and anywhere. Additionally, the National Institute of Standards and Technology (NIST) contributes to a more technical definition of the term, in which they identified cloud computing as a model that enables ubiquitous, practical, on-demand access to cloud resources that can be offered and released with minimal administration effort or engagement from cloud service providers [12].

According to NIST [12], a true cloud solution can be validated based on five basic characteristics of cloud computing, which are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measure service. Currently, there are three service models that are prominent in the industry, namely Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Each model is inextricably linked to one another, building a three-tier cloud service which ultimately forms cloud computing (see Fig. 1).

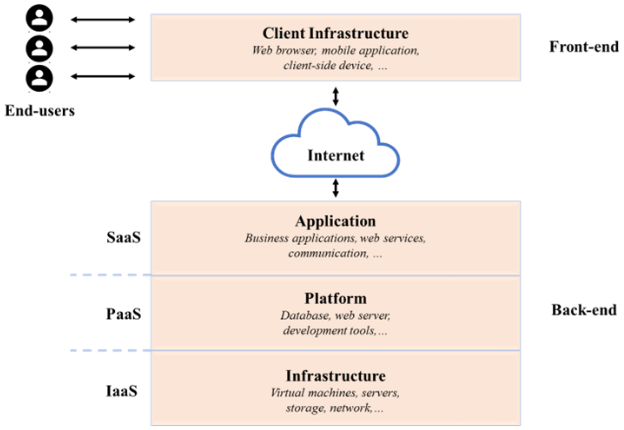

Client Infrastructure
Web browser, mobile application, client-side device, ...
Front-end
End-users
Internet
SaaS
Application
Business applications, web services, communication, ...
PaaS
Platform
Database, web server, development tools, ...
Back-end
IaaS
Infrastructure
Virtual machines, servers, storage, network, ...

Fig. 1. Cloud computing architecture.

In the SaaS model, the applications are hosted by the cloud service providers and offered to the end users over the internet. It means that instead of installing the applications locally on the end user’s computer, he/she may access the applications via the Internet [13]. The end users will only have control over the application settings, while the cloud infrastructure is fully managed by the cloud service providers. SaaS has multi-tenant architecture where it supports multiple tenants sharing a common infrastructure and the model offers services based on pay-per-use [14]. The PaaS model does not only provide a virtualized platform for the end users (i.e.: developers and deployers) to develop and deploy applications on the cloud, but it also offers database services [13]. The end users may use the tools, programming environments and configuration management tools that are provided by the cloud service providers based on a pay-per-use basis. This model can relieve developers of most of the system administration effort (e.g.: setting up and switching between development, test, and production environments), providing flexibility and scalability in PaaS [15]. The IaaS model consists of the hardware layer (e.g.: control processing unit (CPU), memory, disk, bandwidth) and the infrastructure layer (e.g.: virtual machine (VM)) that holds the servers, network and operating system provisioned by the cloud service providers [13]. The distinct feature of IaaS is scalability, also known as on-demand scalability, where the cloud infrastructure is rented as virtual machines based on a pay-per-use manner that dynamically scales in and out based on customers’ demands [16].

In cloud computing environment, the cloud acts as a virtual computing environment with different options to deploy cloud services depending on the business needs. NIST listed four main deployment models which are public cloud, private cloud, community cloud, and hybrid cloud [12]. The public cloud offers general and public access to the cloud services and due to its open access, the security challenges for this model are critical because the resources are shared among multiple cloud consumers [17]. One simple example of a public cloud service is Google Drive that offers storage spaces located on the cloud for public users to access anytime and anywhere with internet access. The private cloud offers services to be deployed within an organization and is treated as an intranet functionality with the billing is subscription bases. Some examples include Amazon Web Services (AWS) Outposts, OpenStack, Microsoft Azure Stack, and others. The community cloud works in a similar way to the private cloud, but the model is exclusive to a group of organizations that share common interests, like compliance policies, security, and mission objectives. For instance, three companies shared the storage between them and therefore, reducing the installation costs if they shared a common infrastructure. Lastly, in the hybrid cloud model, the cloud infrastructure is set up of two or more types of cloud models (private, public, or community), each of which remains a separate legal entity, but are connected by standardized technology that enables the portability of data and applications [12].

In addition to the previous four deployment models, multi-cloud is a deployment model that deploys cloud services on multiple clouds and uses multiple cloud providers. One of the benefits of this deployment model is it allows redundancy, where resources are made available on different platforms to prevent data loss or a malware attack [18]. This study will be focusing on multi-cloud deployment model that will be explained further in the next section.

B. Multi-Cloud Computing

Multi-cloud computing refers to the adoption of multiple independent cloud architectures that act as a single cloud architecture, where applications are distributed across these clouds in discrete pieces [2]. As opposed to hybrid clouds, the components of a multi-cloud system are all distinct cloud systems rather than deployment models [19]. For example, a multi-cloud system is composed of two or more public clouds (see Fig. 2(a)), while a hybrid cloud system is a combination of a public cloud and a private cloud (see Fig. 2(b)).

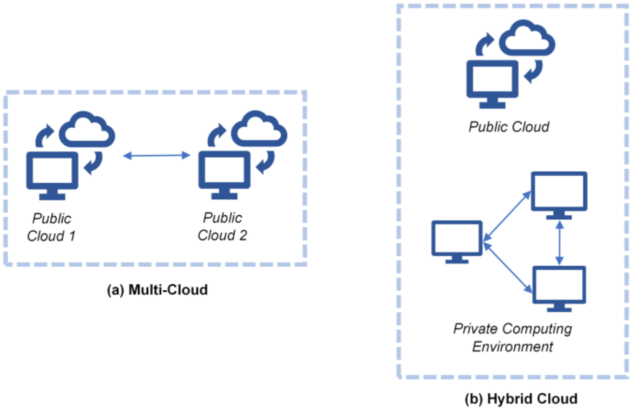


Fig. 2. Multi-cloud vs hybrid cloud deployment model.

Varghese and Buyya [6] emphasized that the changes in cloud computing environment are inevitable, and this leads to the changes of cloud infrastructure. Many existing cloud applications are hosted on data centers of a single provider, and thus creating several challenges like high energy consumption by a single large data center, centralized cloud data centers are susceptible to single point failures, and more. Therefore, they suggested by adopting the multi-cloud strategy these challenges can be overcome. However, due to the heterogeneous nature of multiple cloud providers, adopting the multi-cloud strategy can be challenging because of problems like different APIs, data formats, networks, and storage architectures across providers. It eventually prevents clouds from becoming interoperable, from exchanging data to migrating applications from one cloud to another. Hence, cloud interoperability in multi-cloud platforms is a critical issue to be handled.

C. Cloud Interoperability

According to IEEE international standard vocabulary, interoperability is referring to the ability of two or more systems, products, or components to communicate information and utilize that information [4]. Therefore, in general, cloud interoperability is the ability of the systems to interoperate across different cloud platforms. Nogueira et al. [20] suggested that the term “cloud interoperability” describes the capacity to create applications that integrate resources from different cloud providers to capitalize the unique features offered by each cloud provider.

Achieving cloud interoperability across multiple platforms is a challenge to overcome due to the distinct offerings, technologies, and interfaces by different cloud providers. Thus, complicates the process of merging or shifting resources and services between these different cloud platforms. The general issue with cloud interoperability is that different providers do not utilize common APIs, which makes it difficult for users and applications to interact with the cloud [6, 7, 8, 9]. Additionally, Varghese and Buyya [6] have listed other common issues like diverse network and storage architectures across different providers, significant programming effort is needed to develop a multi-cloud application due to the format differences of multiple providers, and manual management of tasks due to the lack of common interfaces. On top of that, Birje et al. [9] added that it is hard to detect the fault in data transmission across applications and clouds. As a result of these issues, consumer’s acceptance and adoption of multi-cloud strategy is hampered [8].

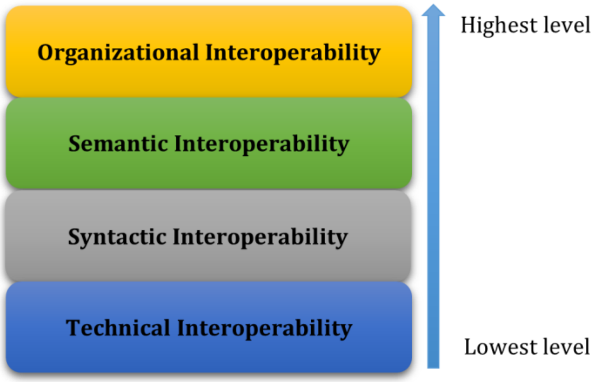

Organizational Interoperability
Semantic Interoperability
Syntactic Interoperability
Technical Interoperability
Highest level
Lowest level

Fig. 3. Levels of interoperability and their correlations.

In general, most interoperability solutions are considered in four levels, with each level signifying a varying degree of compatibility and integration across cloud systems [10, 21, 22] (see Fig. 3). By having these interoperability levels, organizations can evaluate their ability to collaborate with various cloud providers and systems.

Based on Fig. 3, the interoperability levels and their correlations are described as the following:

1) Technical interoperability: This level is considered the lowest level of interoperability because it focuses on the technological facets of interoperability. For example, enabling the exchange of data and communication between multiple cloud systems across different platforms and infrastructures [10]. It covers technical aspects like interface specifications, data integration services, secure communication protocols, and data presentation [21].

2) Syntactic interoperability: In this interoperability level, it is often paired with the implementation of technical interoperability. This level is concerned with standardization of data formats to allow data to be exchanged among cloud systems [22]. It can be done by specifying the exact syntax and format of the data to be exchanged.

3) Semantic interoperability: This level is essential in interoperability between different systems as it tries to ensure that the meaning of the exchanged data can be understood and interpreted correctly [10]. In this level, users must share a common understanding of the data, metadata, and procedures used by various cloud platforms. To make data exchange and processing easier, standardized data models, ontologies, and metadata standards are frequently needed [22].

4) Organizational interoperability: This level addresses interoperability at a higher level of abstraction where it includes coordinating business processes, workflows, and automation across multiple cloud environments [21]. The interoperability at this level is dependent on the successful implementation of the previous interoperability levels: technical, syntactical, and semantic interoperability [22].

This study will be focusing on the semantic interoperability level in multi-cloud platforms. Multi-cloud strategy has developed as a strategic solution in the changing landscape of modern Information Technology (IT) infrastructure, enabling organizations to optimize performance, resilience, and cost-effectiveness by distributing their workloads over several cloud service providers. This paradigm shift reduces the risk of vendor lock-in and increases overall flexibility by allowing businesses to leverage the strengths of multiple cloud platforms. Thus, enabling semantic interoperability in a multi-cloud context is critical because semantic interoperability can improve application and data portability [23].

The term “semantic” is about understanding and making sense of words [24]. Semantic interoperability can be defined as the ability to exchange data in a meaningful way between two or more systems [25]. Semantic interoperability in multi-cloud platforms refers to the ability of disparate cloud systems and services to communicate seamlessly by having a mutual understanding of exchanged data and thus being able to use the data in a meaningful way. In a scenario where organizations opted for multiple cloud providers and services to meet their computing and storage needs, data exchange and processing can be a challenge due to unique data formats, structures, and policies offered by these providers and services. Semantic interoperability can address this challenge by guaranteeing that data and information can be exchanged effortlessly and understood across diverse cloud platforms without ambiguity and loss of meaning.

However, establishing semantic interoperability in the context of multi-cloud platforms has various challenges and limitations. At present, there are no industry-wide standards for semantic representation among various cloud providers [26]. But rather, most research activities and existing standard-setting entities produce various standardization efforts by resolving semantic interoperability problems from multiple perspectives. Another challenge of semantic interoperability is ensuring that performance, scalability, security, and other metrics are not compromised as these metrics are considered the quality attributes of any cloud systems [27]. For example, as the size and complexity of multi-cloud environments increase, ensuring semantic interoperability at scale becomes increasingly important because cloud computing services are scalable to meet the needs of the consumers. On top of that, implementing semantic interoperability is a complex task due to potential data model and ontology mismatches that can only be discovered during runtime [28]. It is due to the wide range of data models and ontologies utilized by various cloud services. Nonetheless, despite the challenges and limitations, research initiatives in semantic interoperability in multi-cloud platforms should continue for a variety of compelling reasons, including meeting evolving business needs, enhancing cross-cloud collaboration, mitigating vendor lock-in, optimizing resource utilization, and others.

III. Extended Cloud Computing Interoperability Taxonomy

A well-defined taxonomy for semantic cloud interoperability is necessary to provide a better understanding of the topic and address the relationships between different elements of a multi-cloud ecosystem. The extended taxonomy is built upon an existing cloud computing interoperability (CCI) taxonomy by Ayachi et al. [10]. There are three main axes in Ayachi et al.’s taxonomy:

- CCI factors: It is comprised of five principal factors of CCI solutions, which are CCI deployment level, CCI interaction patterns, CCI consumer-centric, CCI approach, and CCI time-line perspective.

- CCI scenarios: It refers to the scenarios match with the use cases that have been studied previously, which includes provider-side scenarios and client-side scenarios.

- CCI solutions: It presents existing research efforts on proposed solutions to enable cloud interoperability.

To extend the existing CCI taxonomy, this study explores four types of approaches for semantic interoperability based on the work in [29, 30], and they are:

1) Semantic approach: The semantic approach primarily focuses on the semantic or meaning of data. In the cloud landscape, the implementation of semantic approach is through semantic web technologies, which includes defining shared ontologies, vocabularies, and semantic models to enable uniform data interpretation across cloud platforms [31]. One of the challenges of this approach is due to distinctive ontologies, it is difficult to align them and ensuring effective semantic mappings between them [32].

2) Standard-based approach: In order to facilitate semantic interoperability when exchanging data, the standard-based approach places a strong emphasis on the establishment of common standards, protocols, and data formats. Several standardization efforts have been made by standardization bodies and organizations that cover standards concerning development, security, management, deployment, and other matters related to the cloud platforms. In the work by Kaur et al. [29], the authors provided a complete list of organizations with their standardization projects. However, the main issue with this approach is no standard has been accepted universally to solve the semantic interoperability problems [26].

3) Model-based approach: The model-based approach centers around developing and deploying shared models that represent data structure and semantics. These models are developed using common modeling languages such as Unified Modeling Language (UML) or domain specific languages. This approach, however, is limited by the ability to transition between models and actual solution implementations [33]. Cloud Modelling Framework (CloudMF), which uses model-driven engineering to facilitate provisioning and deploying multi-cloud applications, is an example of a model-based approach [34].

4) Open libraries and open services: This approach relies on the use of abstraction layers and adapters, which support interoperability in the context of multi-cloud platforms. Some of the well-known open libraries are Apache jclouds and Apache Libcloud. For open services, some examples include RightScale, enStratius, and Kaavo [29].

This study will further investigate the semantic approach to fulfil the study's goal. Recent existing literature was studied with an aim to identify the best semantic approach for facilitating semantic interoperable clouds. In a survey by Adhoni [23], the author studies three popular approaches for building semantic interoperability solutions: semantics, frameworks, and standards. He claims that common APIs and data models are key solutions in semantic approach. DiMartino et al. [35] provide three categories of cloud portability and interoperability solutions: framework and model-based approaches, adapting methodologies, and standardization efforts. They stated that using semantic modeling can be beneficial in three aspects of cloud computing: to define the functionalities of applications and quality-of-service details regardless of the platforms, creating models for representing metadata, and enhancing service descriptions between different platforms. Additionally, semantic web technologies like Web Ontology Language (OWL), Web Ontology Language for Service (OWL-S), Resource Description Framework (RDF), SPARQL Protocol and RDF Query Language (SPARQL), and The Semantic Web Rule Language (SWRL) can be used to address these three aspects. According to Kaur et al. [29], the two approaches that are typically recommended for achieving semantic interoperability are standardized APIs and data models. In addition to the two approaches already stated, the authors noted from existing research that using a broker can help ease semantic interoperability by making it easier for users to match their needs with those of cloud vendors. In a systematic review by Tomarchio et al. [30], the authors have concluded their review with four interoperability approaches: open standards, semantics, model-driven engineering (MDE), and open libraries and services. Under the semantics approach, the authors claimed that employing semantic technologies (e.g.: OWL, SPARQL, and SWRL) for achieving semantic interoperability is a proven solution. Bouzerzour et al. [36] discovers that the most adopted approach for achieving cloud service interoperability is the use of semantic technologies, which is from the client's perspective. They have identified existing works that use APIs, ontology, semantic engine, inference rules, and semantic annotation to achieve semantic interoperability. As per Ramalingam and Mohan [26], the semantic level for portability and interoperability of cloud services can be addressed with semantic cloud ontology (e.g.: OWL and OWL-S) and frameworks. The authors have reviewed existing efforts on interoperable and portable frameworks based on three semantic technologies (i.e.: OWL, WSDL, and RDF) and discovered that the representation of semantic cloud services and resources lacks a common or uniform approach.

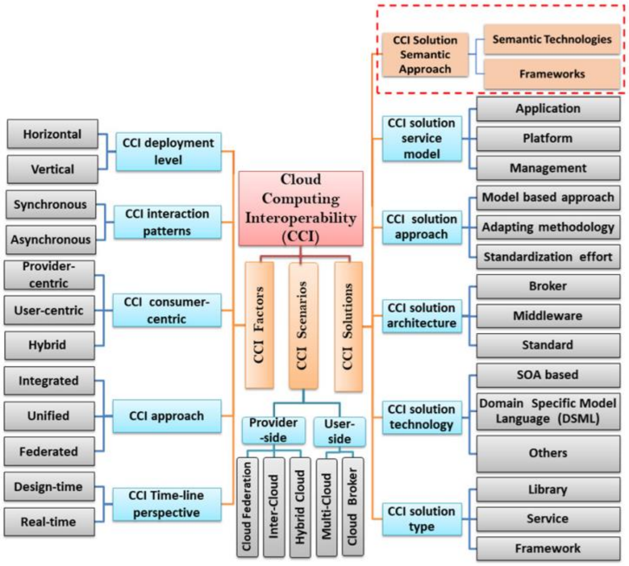


Fig. 4. Extended taxonomy for cloud computing interoperability.

As a result, considering the discussions in the preceding paragraph, this study suggests extending the existing cloud computing interoperability taxonomy by Ayachi et al. [10] by adding a sixth CCI Solution, namely CCI Solution Semantic Approach, with its two approaches which are Semantic Technologies and Frameworks. As shown in Fig. 4, they are depicted within a dashed rectangle. The details of CCI Solution Semantic Approach are discussed in the next subsections.

B. Semantic Technologies

In recent years, semantic technologies were frequently used to accomplish semantic cloud interoperability by providing a common platform for understanding and representing data and services in the cloud. Semantic technologies or sometimes called semantic web technologies comprised of a set of methods, tools, and standards that specifically deals with the meaning and interpretation of data to be understood by machines and applications [36]. Once data is interpreted correctly, these machines and applications can process the data more intelligently.

Current literature agrees that ontologies are the core element of semantic solutions [36, 26, 37, 38]. An ontology reflects domain knowledge, where the ontology classes are typically depicted using graphical models, as models are considered to have explicit semantics [39]. Ontologies are often expressed in a formal language that can be interpreted by both humans and machines, like RDF and OWL. Applications such as information retrieval, semantic web, natural language processing, data integration, and knowledge management all heavily rely on ontologies. They provide a shared understanding among various systems and users by offering an organized and standardized means of representing and exchanging knowledge. Al-Sayed et al. [40] stated that three fundamental components build up an ontology: classes (or concepts), objects (or instances), and properties (or relations). The class is used to describe a collection of instances with comparable characteristics. Properties are used to indicate relationships between instances (i.e.: object property) or between instances and data (i.e.: data-type property). Besides ontologies, other semantic technologies that have been employed for semantic cloud interoperability are semantic APIs, semantic engine, inference rules, and semantic annotations [36].

One of the recent works on ontologies for semantic interoperability is MIDAS-OWL, which is an ontology built on OWL to formally represent the interactions between SaaS and Data as a Service (DaaS) [41]. The proposed ontology connects data among DaaS by rewriting queries with semantically identical properties. PaaSport semantic model is another OWL-based ontology to best support an algorithm for semantic matchmaking and ranking, which suggests the most appropriate PaaS offerings to the application developer [42].

C. Frameworks

According to Partelow [43], the term "framework" has multiple definitions and purposes depending on the field in which it is used. Hence, the author provides several notable definitions of the term “framework” in different contexts and fields of study. It can be concluded that a framework is a methodical and well-structured collection of ideas, procedures, and tools that serves as a basis for creating and addressing complex problems. In software development, for example, a framework provides scaffolding for guiding the overall design and implementation of a system or a project. In addition to providing a pre-established structure and design patterns, frameworks frequently come with tools and libraries that facilitate the development process. Therefore, it offers several benefits to software development by helping to speed up the development process, ensure consistency across projects, promotes flexibility and easy to reuse the components.

Due to businesses’ growing interest in multi-cloud strategy and the need to ensure that the clouds are semantically interoperable, the solutions for semantically interoperable clouds demand an efficient framework. Frameworks that support semantic cloud interoperability, either through standardized interfaces or protocols, can be useful when implementing solutions within the context of a reference architecture. This is because reference architecture acts as a blueprint or template for the design and development of concrete architectures in IT domains. It is considered as a one-to-many relationship between a particular implementation and concrete architectures, and thus, it is an abstract representation of that architecture [44]. It outlines the recommended components, interfaces, and protocols to enable seamless interaction and integration between cloud platforms. It also includes guidelines and best practices for security, scalability, performance, portability, interoperability, and management [45]. Furthermore, Valle et al. [46] revealed that existing works did not explicitly propose techniques (e.g.: models, procedures, or other terminology) to characterize interoperability in reference architectures. As such, the authors emphasized the need to suggest novel approaches for modelling interoperability in reference architecture and for addressing interoperability during architecture instantiation. Therefore, this study suggests that building a cloud solution using a reference architecture is seen as a fitting approach that can contribute to a uniform representation of semantic-based solutions.

Existing prominent organizations and interest groups have produced their own reference architectures or open frameworks and made them freely available for generating further innovative solutions. These architectures or frameworks are offered as open standards that acknowledge the various needs for heterogeneous ecosystems and have been set as a rule for cloud interoperability. Bakshi and Beser [47] review eleven existing reference architectures from standards bodies, consortiums, and forums. In their review, they emphasized that the NIST Cloud Computing Reference Architecture (CCRA) has generic cloud computing architectural building blocks with five major actors (i.e.: Cloud Consumer, Cloud Provider, Cloud Carrier, Cloud Auditor, and Cloud Broker). Each actor has its own roles that are important for managing and providing cloud services in the cloud [45]. Other reference architectures are concerned more on the compliance towards standards, networking and communication services, security, cloud infrastructure, cloud management systems, and cloud interoperability.

Sana et al. [48] review nine existing reference architectures by NIST, Oracle, Distributed Management Task Force

(DMTF), International Business Machines Corporation (IBM), Hewlett-Packard (HP), Cisco, Cloud Security Alliance (CSA), Storage Networking Industry Association (SNIA), and Elastra. The authors concluded with three categories of reference architectures in cloud computing, and they are:

1) Role-based: In this form of reference architecture, the services and activities are matched to roles like cloud service providers and cloud consumers. The architectures of this kind of categories are NIST, Oracle, and DMTF.

2) Layer-based: This form of reference architecture maps services and activities to various architectural layers, including resource layers, application layers, service management layers, and security layers. The architectures of this kind of categories are IBM, HP, and Cisco.

3) Context-based: Reference architectures of this kind offer specific configurations to suit customer needs and make their adoption easier. The architectures of this kind of categories are CSA, SNIA, and Elastra.

Therefore, based on the review by Bakshi & Beser [47] and Sana et al. [48], out of all the reference architectures, the NIST CCRA and IBM CCRA can easily be adapted to this study. These two CCRAs can be a good reference and guideline as they systematically represent the reference architecture using role-based and layer-based architectures. Hence, a semantic interoperability layer can be added as part of the proposed reference architecture for semantic multi-cloud interoperability.

IV. Related Works and Discussion

Even though several cloud computing reference architectures have been produced by notable organizations and interest groups (e.g.: NIST, IBM, AWS, and Google Cloud), the main purpose of these reference architectures is to establish common frameworks and guidelines for industry-wide adoption. Typically, standard-based architectures are more stringent and directive to guarantee uniformity and compliance amongst implementations [49]. As a result, reference architectures have been produced by research initiatives to broaden the possibilities available to other researchers and industries. Research-driven reference architectures are often developed to explore new and test ideas, concepts, or technologies. They may not necessarily aim for immediate standardization.

Therefore, this study aims at reviewing recent (2018-2023) research-driven reference architectures for achieving semantic multi-cloud interoperability. Thirteen recent related works on semantic interoperability in multi-cloud platforms are selected, while the works in other domains such as the Internet-of-Things (IoT), Artificial Intelligence (AI), Blockchain, and big data are excluded. The review focuses on the authors' contributions and summarizes the findings based on the six CCI Solutions (as depicted in the extended cloud computing interoperability taxonomy in Fig. 4). The findings are reported in Table I.

As shown in Table I, it is found that most of the semantic approaches are using ontologies, like OWL, OWL-S, and RDF. This indicates that ontologies are proven solutions for semantic interoperability. Other than ontologies, APIs are among the popular solutions as they can serve as common interfaces between different platforms. The solutions are implemented as models, common architectures, and even toolkits. For the models, most works proposed semantic models that can be employed as part of a framework, stored as libraries, and adopted in semantic engine or semantic layer. For the common architectures, they are represented using layer-based, role-based, and hybrid of both. Apart from that, not every proposed solution aims to address interoperability across the three cloud service models (i.e.: IaaS, PaaS, and SaaS). Most solutions address interoperability independently for each of the three cloud service models. Existing works also prefer providing solutions based on broker and middleware architecture. Standards are the least preferred in developing research-driven architectures. SOA based are seen as the most favorable implementation of the solutions. It might be because broker and middleware architectures can effectively complement the SOA based implementation. The type of solutions produced is mostly frameworks and libraries. Finally, the limitations of each reviewed work are presented in the last column of Table I.

TABLE I. Review of Recent Related Works

Related Works (2018-2023)	**CCI Solutions**						**Limitations**
	CCI Solution Semantic Approach	***CCI Solution Service Model***	***CCI Solution Approach***	***CCI Solution Architecture***	***CCI Solution Technology***	***CCI Solution Type***	
FCLOUDS framework to achieve semantic interoperability in multi-clouds [50]	API, Open Cloud Computing Interface (OCCI)	Application	Model based approach	Middleware	DSML	Framework	Limited to the OCCI standard.
A common interoperable model for cloud computing [48]	API, Architecture is a hybrid of role-based & layer-based	Application, Platform, Management	Adapting methodology	Standard	SOA based	Framework	API is insufficient for semantic understanding.
PaaSport semantic model: An ontology for a platform-as-a-service semantically interoperable marketplace [42]	OWL, RDF, API, Service Level Agreement (SLA), Semantic model (layer-based)	Platform	Model based approach	Broker	SOA based	Service	The solution is specific to PaaS offerings.
CloudLightning Ontology (CL-	OWL, Semantic engine	Platform, Management	Model based approach	Middleware	SOA based	Library	Currently, the system cannot evaluate

Related Works (2018-2023)	**CCI Solutions**						**Limitations**
	CCI Solution Semantic Approach	***CCI Solution Service Model***	***CCI Solution Approach***	***CCI Solution Architecture***	***CCI Solution Technology***	***CCI Solution Type***	
Ontology): An ontology for heterogeneous resources management interoperability and HPC in the cloud [51]							performance, resource utilization, and energy consumption.
PacificClouds: A flexible microservices based architecture for interoperability in multi-cloud environments [52]	API, SLA, Microservice	Management	Adapting methodology	Broker	SOA based	Service	API is insufficient for semantic understanding. The implementation of semantic SLA is not clearly stated.
EasyCloud: A rule-based toolkit for multi-platform Cloud/Edge service management [53]	API, toolkit	Application	Adapting methodology	Broker	SOA based	Library	API is insufficient for semantic understanding. The solution is specific to SaaS offerings.
Cloud interoperability based on a generic cloud service description: Mapping OWL-S to GCSD [54]	Mapping rules (OWL-S), Pivot model (mediator for transforming different cloud service description languages to the GCSD).	Application.	Model-based approach.	Middleware.	DSML	Library	The solution is specific to SaaS offerings.
EasyCloud toolkit to effectively support the creation and usage of Multi-cloud Systems (MSs) [55]	API, toolkit	Application	Adapting methodology	Broke	SOA based	Library	API is insufficient for semantic understanding. The solution is specific to SaaS offerings.
Semantic Interoperability Framework for IAAS Resources in Multi-Cloud Environment [56]	RDF, OWL, SPARQL, Ontology mapping	Management	Model-based approach	Broker	SOA based	Framework	The solution is specific to IaaS resource management.
MIDAS: A domain specific language to provide middleware for interoperability among SaaS and DaaS/ DBaaS through a metamodel approach [57]	API, Semantic mapping, Structured Query Language (SQL) or Not Only SQL (NoSQL), Metamodel (Eclipse Modeling Framework (EMF)	Application	Model-based approach	Middleware	DSML	Framework	Limited to Middleware for DaaS/DBaaS and SaaS (MIDAS) architecture.
Cloud Enterprise Resource Planning (ERP) API Ontology [58]	OWL	Application, Platform, Management	Model-based approach	Middleware, Broker	SOA based	Framework	No ontology evaluation in a practical application for business data migration between cloud ERP providers.
PaaS and IaaS Resource Semantic Interoperability Framework (extended from their previous work) [59]	RDF, OWL, SPARQL, Ontology mapping	Platform, Management	Model-based approach	Broker	SOA based	Framework	The solution is specific to PaaS and IaaS resource management.
Cloud Interoperability Pivot Model (CIPiMo) for cloud service interoperability [60]	Mapping rules (WSDL and OWL-S), Pivot model (mediator for transforming different cloud service description languages to the GCSD)	Application	Model-based approach	Middleware	DSML	Library	The solution is specific to SaaS offerings.

Even though several research efforts have been done related to the topic of this study, it is found that existing works on reference architectures for semantic multi-cloud interoperability are still not in a mature stage and prompting for future works. Some research gaps that can be identified from the review are as the following:

- The integration of semantic-based solutions (e.g.: semantic model) within a framework (e.g.: reference architectures) is not explicitly and uniformly represented. This is an essential consideration in current research given the growing interest in multi-cloud strategy and the fact that multi-cloud platforms are inherently diverse.
- Cloud interoperability solutions are not inclusive of three cloud service models (i.e.: IaaS, PaaS, and SaaS). Most of the works address interoperability independently across the three cloud service models. It is important to consider cloud interoperability vertically and horizontally across the three cloud service models.

By identifying the research gaps in recent related works, this study aims to propose future works that attempts to solve these problems. Therefore, the following are suggested for future works of this study:

- To identify necessary requirements for developing ontologies and reference architectures.
- To develop a semantic model utilizing ontologies and other semantic technologies in order to facilitate semantic interoperability in data exchange across various cloud platforms.
- To develop a reference architecture for semantic multi-cloud interoperability by adapting a hybrid of role-based and layer-based architectures with an extended semantic interoperability layer. In addition to that, there is a need to identify the required roles (actors) and layers for the proposed reference architecture.
- To implement the proposed semantic model and reference architecture against use cases in multi-cloud platforms.

V. Conclusion

Semantic interoperability in multi-cloud platforms enables uniform interaction and interpretation of data and applications across diverse platforms. However, achieving semantic interoperability remains a challenge, and research efforts in this area are ongoing. Although professional groups or organizations have developed standard solutions for semantically interoperable clouds, research-driven initiatives are still required to enhance semantic interoperability by delivering uniform blueprints.

This study explores the significance of multi-cloud computing in today’s dynamic and complex IT landscape. As organizations increasingly rely on cloud services for their diverse computing needs, understanding the implications and benefits of adopting a multi-cloud strategy becomes paramount. The study investigates how multi-cloud environments can address the evolving requirements of scalability, flexibility, and other requirements while navigating interoperability concerns. By examining the recent related works on two main areas, which are semantic interoperability and reference architecture, this study aims to provide a strategic solution in the form of a reference architecture for semantic interoperability in multi-cloud platforms. Furthermore, the review on recent related works reveals that the lack of widely accepted semantic models and frameworks indicates that the field of study is still in its infancy and needs further development.

Therefore, two contributions have been proposed in this study, and they are:

- An extended CCI taxonomy by adding the semantic approach which consists of semantic technologies and frameworks that are considered crucial approach to enable an effective semantic interoperability in multi-cloud platforms. This taxonomy can serve as a knowledge base for future researchers and promote consistency across different research studies.
- A review of recent related works on semantic interoperability in multi-cloud platforms, highlighting the current CCI solutions employed by the authors in their proposed work. The review includes the limitations of each work, and thus prompting for future work. The result of this review is not only important for studying the current technologies used for semantic interoperability, but also for identifying the research gaps that may present in current research.

References

[1] S. Mathew and J. Varia, "Overview of Amazon Web Services AWS whitepaper," Amazon Web Services, Seattle, WA, 2020.

[2] I. Odun-Ayo, M. Ananya, F. Agono and R. Goddy-Worlu, "Cloud computing architecture: A critical analysis," in 18th International Conference on Computational Science and Its Applications, ICCSA 2018, Melbourne, 2018.

[3] A. Krishna, S. Cowley, S. Singh and L. Kesterson-Townes, "Assembling your cloud orchestra: A field guide to multicloud management.," 2018. [Online]. Available: https://www.ibm.com/thought-leadership/institute-business-value/report/multicloud.

[4] ISO, "IEC/IEEE International Standard-Systems and Software Engineering–Vocabulary," ISO/IEC/IEEE 24765: 2017 (E), 2017.

[5] J. Opara-Martins, R. Sahandi and F. Tian, "Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective," Journal of Cloud Computing, vol. 5, no. 1, pp. 1-18, 2016.

[6] B. Varghese and R. Buyya, "Next generation cloud computing: New trends and research directions," Future Generation Computer Systems, vol. 79, pp. 849-861, 2018.

[7] Y. Al-Dhuraibi, F. Paraiso, N. Djarallah and P. Merle, "Elasticity in cloud computing: state of the art and research challenges," IEEE Transactions on Services Computing, vol. 11, no. 2, pp. 430-447, 2017.

[8] G. T. Ayem, S. G. Thandekkattu and N. R. Vajjhala, "Review of interoperability issues influencing acceptance and adoption of cloud computing technology by consumers," in Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021, Singapore, 2022.

[9] M. N. Birje, P. S. Challagidad, R. H. Goudar and M. T. Tapale, "Cloud computing review: concepts, technology, challenges and security," International Journal of Cloud Computing, vol. 6, no. 1, pp. 32-57, 2017.

[10] M. Ayachi, H. Nacer and H. Slimani, "Cloud computing interoperability: An overview," in 2nd International Conference on New Technologies of Information and Communication, NTIC 2022, Virtual, Online, 2022.

[11] J. S. Hurwitz and D. Kirsch, Cloud computing for dummies, John Wiley & Sons, 2020.

[12] P. Mell and T. Grance, "The NIST definition of cloud computing, Special Publication (NIST SP)," National Institute of Standards and Technology, Gaithersburg, MD, 2011.

[13] S. Namasudra, "Cloud computing: A new era," Journal of Fundamental and Applied Sciences, vol. 10, no. 2, 2018.

[14] S. Liu, K. Yue, H. Yang, L. Liu, X. Duan and T. Guo, "The research on SaaS model based on cloud computing," in 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2018, Xi'an, 2018.

[15] A. Rashid and A. Chaturvedi, "Cloud computing characteristics and services: A brief review," International Journal of Computer Sciences and Engineering, vol. 7, no. 2, pp. 421-426, 2019.

[16] E. B. Chawki, A. Ahmed and T. Zakariae, "IaaS cloud model security issues on behalf cloud provider and user security behaviors," in Procedia Computer Science, Las Palmas de Gran Canaria, 2018.

[17] G. Ramachandra, M. Iftikhar and F. A. Khan, "A comprehensive survey on security in cloud computing," in 14th International Conference on Mobile Systems and Pervasive Computing, MobiSPC 2017, Leuven, 2017.

[18] P. Wang, C. Zhao, W. Liu, Z. Chen and Z. Zhang, "Optimizing data placement for cost effective and high available multi-cloud storage," Computing and Informatics, vol. 39, no. 1-2, pp. 51-82, 2020.

[19] J. Hong, T. Dreibholz, J. A. Schenkel and J. A. Hu, "An overview of multi-cloud computing," in 33rd International Conference on Advanced Information Networking and Applications, AINA 2019, Matsue, 2019.

[20] E. Nogueira, A. Moreira, D. Lucrédio, V. Garcia and R. Fortes, "Issues on developing interoperable cloud applications: definitions, concepts, approaches, requirements, characteristics and evaluation models," Journal of Software Engineering Research and Development, vol. 4, no. 1, pp. 1-23, 2016.

[21] M. Kostoska, M. Gusev and S. Ristov, "An overview of cloud interoperability," in Federated Conference on Computer Science and Information Systems, FedCSIS 2016, Gdansk, 2016.

[22] P. H. D. Valle, L. Garcés and E. Y. Nakagawa, "A typology of architectural strategies for interoperability," in 13th Brazilian Symposium on Software Components, Architectures, and Reuse, SBCARS 2019, Salvador, 2019.

[23] Z. A. Adhoni, "Framework, semantic and standard approaches in multi-clouds to achieve interoperability: A survey," Journal of Integrated Science and Technology, vol. 10, no. 2, pp. 67-72, 2022.

[24] L. Rachana and S. Shridevi, "A literature survey: Semantic technology approach in machine learning," in 2nd International Conference on Power Engineering, Computing and Control, PECCON 2019, Chennai, 2021.

[25] M. Sreenivasan and A. M. Chacko, "Interoperability issues in EHR systems: Research directions," in Data Analytics in Biomedical Engineering and Healthcare, Elsevier, 2020, p. 13–28.

[26] C. Ramalingam and P. Mohan, "Addressing semantics standards for cloud portability and interoperability in multi cloud environment," Symmetry, vol. 13, no. 2, pp. 1-18, 2021.

[27] H. A. Imran, U. Latif, A. A. Ikram, M. Ehsan, A. J. Ikram, W. A. Khan and S. Wazir, "Multi-cloud: A comprehensive review," in 23rd IEEE International Multi-Topic Conference, INMIC 2020, Bahawalpur, 2020.

[28] A. Bergmayr, U. Breitenbücher, N. Ferry, A. Rossini, A. Solberg, M. Wimmer, G. Kappel and F. Leymann, "A systematic review of cloud modeling languages," ACM Computing Surveys, vol. 51, no. 1, pp. 1-38, 2018.

[29] K. Kaur, D. S. Sharma and D. K. S. Kahlon, "Interoperability and portability approaches in inter-connected clouds: A review," ACM Computing Surveys, vol. 50, no. 4, pp. 1-40, 2017.

[30] O. Tomarchio, D. Calcaterra and G. D. Modica, "Cloud resource orchestration in the multi-cloud landscape: a systematic review of existing frameworks," Journal of Cloud Computing, vol. 9, no. 1, pp. 1-24, 2020.

[31] H. Brabra, A. Mtibaa, L. Sliman, W. Gaaloul and F. Gargouri, "Semantic web technologies in cloud computing: a systematic literature review.," in In 2016 IEEE International Conference on Services Computing (SCC), 2016.

[32] I. Harrow, R. Balakrishnan, E. Jimenez-Ruiz, S. Jupp, J. Lomax, J. Reed, M. Romacker, C. Senger, A. Splendiani, J. Wilson and P. Woollard, "Ontology mapping for semantically enabled applications," Drug discovery today, vol. 24, no. 10, pp. 2068-2075, 2019.

[33] G. Zacharewicz, S. Diallo, Y. Ducq, C. Agostinho, R. Jardim-Goncalves, H. Bazoun, Z. Wang and G. Doumeingts, "Model-based approaches for interoperability of next generation enterprise information systems: state of the art and future challenges," Information Systems and e-Business Management, vol. 15, no. 2, pp. 229-256, 2017.

[34] N. Ferry, F. Chauvel, H. Song, A. Rossini, M. Lushpenko and A. Solberg, "CloudMF: Model-driven management of multi-cloud applications," ACM Transactions on Internet Technology (TOIT), pp. 1-24, 2018.

[35] B. Di Martino, G. Cretella and A. Esposito, "Cloud portability and interoperability," in Encyclopedia of Cloud Computing, 2016, pp. 163-177.

[36] N. E. H. Bouzerzour, S. Ghazouani and Y. Slimani, "A survey on the service interoperability in cloud computing: Client-centric and provider-centric perspectives," Software - Practice and Experience, vol. 50, no. 7, pp. 1025 - 1060, 2020.

[37] A. Patel and S. Jain, "Present and future of semantic web technologies: A research statement," International Journal of Computers and Applications, vol. 43, no. 5, pp. 413-422, 2021.

[38] A. Rejeb, J. Keogh, W. Martindale, D. Dooley, E. Smart, S. Simske, S. Wamba, J. Breslin, K. Bandara, S. Thakur and K. e. a. Liu, "Charting past, present, and future research in the semantic web and interoperability," Future internet, vol. 14, no. 6, p. 161, 2022.

[39] J. Agbaegbu, O. T. Arogundade, S. Misra and R. Damaševičius, "Ontologies in cloud computing—review and future directions," Future Internet, vol. 13, no. 12, p. 302, 2021.

[40] M. M. Al-Sayed, H. A. Hassan and F. A. Omara, "Towards evaluation of cloud ontologies," Journal of Parallel and Distributed Computing, vol. 126, pp. 82-106, 2019.

[41] E. L. F. Ribeiro, M. Souza and D. B. Claro, "MIDAS-OWL: An Ontology for Interoperability between Data and Service Cloud Layers," in 17th Brazilian Symposium on Information Systems: Intelligent and Ubiquitous Information Systems: New Challenges and Opportunities, 2021.

[42] N. Bassiliades, M. Symeonidis, P. Gouvas, E. Kontopoulos, G. Meditskos and I. Vlahavas, "PaaSport semantic model: An ontology for a platform-as-a-service semantically interoperable marketplace," Data and Knowledge Engineering, vol. 113, pp. 81-115, 2018.

[43] S. Partelow, "What is a framework? Understanding their purpose, value, development and use," Journal of Environmental Studies and Sciences, vol. 13, no. 4, p. 510–519, 2023.

[44] J. Soldatos, E. Troiano, P. Kranas and A. Mamelli, "A reference architecture model for big data systems in the finance sector," in Big Data and Artificial Intelligence in Digital Finance: Increasing Personalization and Trust in Digital Finance using Big Data and AI, Springer, Cham, 2022, pp. 3-28.

[45] F. Liu, J. Tong, J. Mao, R. Bohn, J. Messina, L. Badger and D. Leaf, "NIST cloud computing reference architecture," NIST special publication 500-292, 2011.

[46] P. H. D. Valle, L. Garcés, T. Volpato, S. Martínez-Fernández and E. Y. Nakagawa, "Towards suitable description of reference architectures," PeerJ Computer Science, vol. 7, pp. 1-36, 2021.

[47] K. Bakshi and L. Beser, "Cloud reference frameworks," in Encyclopedia of Cloud Computing, Wiley, 2016, pp. 71-88.

[48] K. Sana, N. A. C. E. R. Hassina and B. B. Kadda, "Towards a reference architecture for interoperable clouds," in 8th International Conference on Electrical and Electronics Engineering, ICEEE 2021, Antalya, 2021.

[49] R. &. P. I. A. Zota, "An Overview of the Most Important Reference Architectures for Cloud Computing," Informatica Economica, vol. 18, no. 4, 2014.

[50] S. Challita, F. Zalila and P. Merle, "Specifying semantic interoperability between heterogeneous cloud resources with the FCLOUDS formal language," in 11th IEEE International Conference on Cloud Computing, CLOUD 2018, San Francisco, 2018.

[51] G. G. Castañé, H. Xiong, D. Dong and J. P. Morrison, "An ontology for heterogeneous resources management interoperability and HPC in the cloud," Future Generation Computer Systems, vol. 88, pp. 373-384, 2018.

[52] J. De Carvalho, F. Trinta and D. Vieira, "PacificClouds: A flexible microservices based architecture for interoperability in multi-cloud environments," in 8th International Conference on Cloud Computing and Services Science, CLOSER 2018, Funchal, Madeira, 2018.

[53] C. Anglano, M. Canonico and M. Guazzone, "EasyCloud: A rule based toolkit for multi-platform cloud/edge service management," in 5th International Conference on Fog and Mobile Edge Computing, FMEC 2020, Paris, 2020.

[54] N. E. H. Bouzerzour, S. Ghazouani and Y. Slimani, "Cloud interoperability based on a generic cloud service description: Mapping OWL-S to GCSD," in 29th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE 2020, Virtual, Bayonne, 2020.

[55] C. Anglano, M. Canonico and M. Guazzone, "EasyCloud: Multi-clouds made easy," in 45th IEEE Annual Computers, Software, and Applications Conference, COMPSAC 2021, Virtual, Online, 2021.

[56] K. Benhssayen and A. Ettalbi, "Semantic interoperability framework for IAAS resources in multi-cloud environment," International Journal of Computer Science and Network Security, vol. 21, no. 2, pp. 1-8, 2021.

[57] B. Mane, A. P. Magalhaes, G. Quinteiro, R. S. P. Maciel and D. B. Claro, "A domain specific language to provide middleware for interoperability among SaaS and DaaS/DBaaS through a metamodel approach," in 23rd International Conference on Enterprise Information Systems, ICEIS 2021, Virtual, Online, 2021.

[58] D. Andročec and R. Picek, "Cloud ERP API ontology," in International Conference on Electrical, Computer and Energy Technologies, ICECET 2022, Prague, 2022.

[59] K. Benhssayen and A. Ettalbi, "An extended framework for semantic interoperability in PaaS and IaaS multi-cloud," in Digital Technologies and Applications: Proceedings of ICDTA’22, Fez, Morocco, 2022.

[60] N. E. H. Bouzerzour and Y. Slimani, "Towards a MaaS Service for Cloud Service Interoperability," in 10th International Conference on Model-Driven Engineering and Software Development, MODELSWARD 2022, Virtual, Online, 2022.

Pancreatic Cancer Detection Through Hyperparameter Tuning and Ensemble Methods

Koteswaramma Dodda, Dr. G. Muneeswari
School of Computer Science and Engineering, VIT-AP University, Amaravati 522237, India

Abstract—**Computing techniques have brought about a significant transformation in the field of medical research. Machine learning techniques have facilitated the analysis of vast amounts of data, modeling of complex scenarios, and the ability to make well-informed decisions. This presents an opportunity to develop reliable and effective medical system implementations, which may include the automatic recognition of uncertain issues related to health. Currently, significant research efforts to be directed towards the prediction of cancer, particularly focusing on addressing the various health complications caused by this disease, which can adversely impact multiple organs within the body. Pancreatic Cancer (PC) stands out as a highly lethal form of tumor, with a rather discouraging global five-year survival rate of approximately 5%. The truth behind the early detection increases the survival rate and it also helps the radiologists to give better treatment to those who are affected at early stages. Creatinine, LYVE1, REG1B, and TFF1 are urine proteomic biomarkers that offer a promising non-invasive and affordable diagnostic technique for detecting pancreatic cancer. In this study, a novel model that combines gridsearchCV technique to search and find the optimal combination of hyperparameters for a random forest classifier. In this research a new ensemble method to enhance the performance for classification of pancreatic cancer and non-cancer by using urinary biomarkers which is collected from Kaggle. The implemented model achieved better results of Accuracy 99.98%, F-1 score 99.98, Precision 99.98, and Recall 99.98.**

Keywords*—Pancreatic cancer; Machine learning (ML); urinary biomarkers; grid search hyper parameter tuning; Random Forest (RF)*

I. Introduction

According to cancer statistics, Pancreatic Cancer (PC) is predicted to overtake all other causes of death globally in 2022 [1]. The pancreas is positioned posterior to the stomach and anterior to the spine, with the liver, spleen, and gallbladder encircling it. Pancreas releases hormones into the digestive tract to regulate blood sugar levels. The pancreas, positioned behind the lower stomach, is responsible for both regulating blood sugar levels through hormone production and releasing digestive enzymes. When cancer forms in the tissue of the pancreas, it is known as pancreatic cancer. This is due to its location in stomach which is becoming a difficult task for clinicians (researchers) to find PC. Most instances of Pancreatic Cancer occur in those over 45 like additional risk condition, specific genetic mutations, chronic pancreatitis, diabetes, and advancing age.

Now-a-days, PC has become more prevalent [2]. At initial stages Pancreatic cancer often remains asymptomatic. Symptoms including icterus back or bellyache pain, decreased appetite, unexplained weight loss, and digestive issues may appear as the condition worsens [3]. Hence, there exist significant variations in the structure and presentation of the pancreas among different individuals. Among the various types of PC, Pancreatic Ductal Adenocarcinoma (PDAC) is leads to dead. Pancreatic Cancer, which currently lacks a cure, ranks is among the most severe diseases with one of the poorest survival rates. PC stands out for having the lowest survival rate nearly five years of any cancer, primarily due to its late-stage diagnosis, which stands at 11% [1]. Because early indicators are rare, pancreatic cancer (PC) is infrequently detected in its early stages [4], [5]. As a result, both surgical interventions and treatments involving chemotherapy and radiation are generally ineffective in combating PC [6]. To enhance knowledge about the risk of pancreatic cancer, medical professionals advise patients to undergo additional diagnostic measures, such as biomarker testing and medical imaging scans [7]. There are number of difficulties are reported for imaging early pancreatic cancer. The expense associated with radiological image screening is considerable, making it an improbable choice for widespread pancreatic cancer (PC) screening.

Consequently, researchers are turning their attention towards the utilization of biomarkers as an initial approach to early PC detection. The rapid advancements in genomic sequencing and its diverse techniques, including proteomics, epigenomics, and transcriptomics, generate vast amounts of multi-omics data. Biomarkers in PC serve critical roles in early identification, prognosis, treatment decision-making, and research, resulting in more effective and customized care for patients suffering from this difficult disease. The essential biomarkers that play a pivotal role in the early identification of PC can be found in bodily fluids, which encompass cyst fluid, pancreatic juice, and bile. However, gathering these fluids necessitates invasive methods like surgery or endoscopy. In contrast, blood stands out as a low-risk, cost-effective, and easily repeatable source of tumor biomarkers [8]. Blood contains a wealth of proteomic biomarkers, including Carbohydrate Antigen 19-9 (CA19-9), and transcriptomic biomarkers like Circulating Micro RNAs (miRNAs), which can be detected through RNA sequencing. [9].

Historically, blood has served as the principal reservoir for these biomarkers, although urine presents itself as a viable alternative biological fluid. It makes it simple to collect a non-invasive sample in large quantities and do repeated measurements. As of right now, there is no reliable biomarker for detecting PDAC at early stage. The single biomarker used in clinical practice, serum CA19-9, is used largely as a prognostic indicator and to track therapy effectiveness because it is neither specific nor sensitive enough for screening [10]. Urine stands as a hopeful substitute body fluid for the exploration of biomarkers. It proves to be an excellent option for broad diagnostic screening due to its non-invasive and cost-effective nature, as patients can readily provide ample samples [11]. Urine includes proteome indicators much like blood does.

A three-protein biomarker panel that was presented by Radon et al. in 2015 [12] suggests that urine samples can be utilized to diagnose people with early-stage PC. LYVE-1, TFF1, and REG1A were investigated putative proteomic biological markers. In a smaller study, scientists investigated that the miRNA in urine is also used for diagnosis of pancreatic ductal adenocarcinoma at early stage. By substituting REG1B for REG1A in 2020, Debernardi et al. [13] discover an additional protein biomarker panel. Furthermore, because of the comparable symptoms in early-stage PDAC cases, they can differentiate between cases of PDAC and benign hepatobiliary disease, which can be challenging. This analysis takes into account four important urine biomarkers: TFF1, LYVE1, REG1B, and creatinine. Creatinine is likely employed to assess function of kidney. Moreover, Lymphatic Vessel Endothelial Hyaluronan Receptor 1 has associations with tumor metastasis, REG1B with pancreatic regeneration, and Trefoil Factor 1 with the regeneration and urinary tract repair. The analysis of computer-aided diagnosis (CAD) systems has seen the emergence of ML and DL techniques, which serve as the foundation for handling a wide range of information about patients, including radiology, pharmacogenetics, and biological markers.

In the healthcare sector, machine learning (ML) has become a game-changing technology that is revolutionizing the way medical data is analyzed, diagnoses are made, and treatments are given. Random forest serves as a widely employed ML algorithm for addressing both Regression and Classification tasks. GridsearchCV, on the other hand, is the method used for fine-tuning hyperparameters to identify the most effective settings for a given model. AdaBoost, as a potent ensemble learning algorithm, harnesses the collective strength of multiple weak learners to construct a resilient predictive model.

The following is the paper organizational structure. In Section II, a number of related works are addressed. The proposed system methodology is examined in Section III. In Section IV the experimental results are discussed. Finally, Section V concludes the proposed research paper.

II. Literature Review

In the study, Lee et al. [14] detected miRNA indicators released into the bloodstream. The developed diagnostic system for PC by choosing 39 miRNA markers using a penalized support vector machine (SVM) is uses a smoothly trimmed absolute deviation. The model's accuracy was 93%, and AUC of 0.98.

Long et al. [15] identified and validated oncogenic indicators of pancreatic cancer at the genome level using data mining and multi-omics data. Random forest (RF) technique was used to build their prediction system because it is a simple methodology. After effectively uncovering the unseen biological information from multi-omics data, Scientists have pinpointed dependable biomarkers for the early observation, prediction, and diagnosis of PC. The suggested Random Forest (RF) model achieved an impressive 96% of efficiency.

Debernardi and colleagues [16] identified diagnostic miRNAs for detecting pancreatic ductal adenocarcinoma (PDAC) based on urine samples at early stage. The discriminatory potential miRNA biomarkers were identified using LR algorithms. The suggested models yielded the most favorable outcomes, demonstrating 83.3% sensitivity, 96.2% specificity, with AUC 0.92.

Exosomes were used by Ko et al. [17] to diagnose PC utilizing machine learning and nanofluidic technologies. In order to assess unrefined clinical samples, they created a multichannel nanofluidic device. Tenth, to aid in the definitive diagnosis of cancer patients, these exosomes are subjected to the linear discriminant analysis (LDA) technique. The AUC for this prediction model's classification of pancreatic tumours against healthy samples was 0.81.

Lee et al. [18] reported the construction of a prediction model for the people who are in advanced stage from population-based research to find at early stage. The diagnostic methodology acknowledged that will aid in educating the medical community about the risk of pancreatic cancer. NHIRD, Taiwan's health insurance database, served as the foundation for this study. Combination of four models is used to create predictive model: voting ensemble, ensemble learning, deep neural networks, and logistic regression (LR). The model’s AUC ranged from 0.71 to 0.76, and its accuracy was between 73 and 75%.

To identify PC at the localized stage, D Agarwal et al. [19] proposed highly sensitive nano biosensors for protease/arginine detection. The hard hierarchical decision structure (HDS) and soft hierarchical decision structure (SDS) beat traditional multi-class classification methods, achieving an accuracy score of 92%.

Thanya et al. [20] employs color conversion and anis lateral filters on pancreatic cancer CT images from a PCCD database. FK-NNE segmentation and features based on histograms are extracted. A DCNN combined with a DBN classifier classifies pancreatic cell tumors as benign or malignant. The accuracy climbed to 99.6%, with a sensitivity of 100% and a specificity of 99.47%.

1D CNN-LSTM model accurately categorizes patients with pancreatic cancer, outperforming competing models in assessment measures by 97%, Karar et al. [21] based on urine biological markers. The study evaluates several risk prediction algorithms in order to create PancRISK, a biomarker-based risk score. Out of 379 patients were categorized into training and testing sets using urine biomarkers. When a number of machine learning algorithms were compared, none of them performed noticeably better than the others. The PancRISK score was derived using the logistic regression model; however, it could be enhanced by include the PDAC biomarker CA19-9. Currently, Blyuss et al. [22] are investigating the utility of PancRISK for non-invasive patient risk assessment in pancreatic cancer.

In a separate investigation, Jiao et al. [23] delved into the connection between ITGA3 expression in the pancreas and the observations and unhealthy features of patients with pancreatic cancer. Their methodology encompassed data mining and the application of Chi-squared tests to evaluate associations. Their findings unveiled a notable elevation in ITGA3 expression in pancreatic cancer patients, showcasing correlations with microanatomy, stage, and recurrence. Most notably, heightened ITGA3 expressions analogous with diminished persistence rate are especially to find cancer at earlier stage.

In their study, Liu and colleagues [24] investigated 11 long non-coding RNAs (lncRNAs) associated with pancreatic cancer and identified plasma ABHD11-AS1 as a probable biomarker for the detection of pancreatic cancer. They observed that combining ABHD11-AS1 with CA199 improved the efficiency of pancreatic cancer diagnosis compared to using ABHD11-AS1 by itself, especially for early tumor detection. These findings play a potentially important role in lncRNAs for the diagnosis of cancer.

Iwano et al. [25] proposed to develop a PDAC-symptomatic model using serum anabolism from Japanese patients. Utilized liquid chromatography/electrospray ionization mass spectrometry, Researchers developed a diagnostic algorithm based on machine learning. By combining primary metabolites and phospholipids, they enhanced the accuracy of cancer diagnosis and the area under the receiver operating characteristic curve. The incorporation of 36 statistically significant metabolites as a combined biomarker led to a significant 97.4% improvement in results. This system offers an efficient screening approach for pancreatic ductal adenocarcinoma (PDAC).

Discovered by Takahashi et al. [26] Long noncoding RNAs (lncRNAs) play a role in the epithelial-mesenchymal transition (EMT) in pancreatic ductal adenocarcinoma (PDAC). A lncRNA, known as HULC, has been identified as a future biomarker for PDAC. HULC was observed to have elevated expression levels, induced by transforming growth factor-β, in both PDAC cells and their extracellular vesicles (EVs). Suppression of HULC led to reduced invasion and migration of PDAC cells by inhibiting the process of epithelial-mesenchymal transition (EMT). The encapsulation of HULC in EVs may hold promise for aiding in the diagnosis of human PDAC. PDAC poses a significant threat as it is corelated with mortality. Current diagnostic tests and prognostic tests are limited due to the lack of reliable biomarkers. The analysis of methylation in cell-free DNA (cfDNA) holds great promise as a non-invasive method for detecting specific patterns related to disease in pre-neoplastic lesions and chronic pancreatitis.

Brancaccio et al. [27] present a review that delves into the benefits and challenges of cfDNA methylation studies, as well as the latest developments in identifying early diagnostic or prognostic biomarkers for Pancreatic Cancer. In the provision of patient care with invasive malignancies, biomarkers are crucial. Pancreatic Cancer with a dismal prognosis is because of its advanced stage and lack of available treatments. The lack of predictive biomarkers and proven screening tools for early diagnosis and targeted therapy exacerbates this. In the last two decades, there has been a growing body of information on biomarkers related to pancreatic cancer. The limited sensitivity and specificity of CA 19-9 have hindered its exclusive approval as the sole biomarker for diagnosis and evaluating treatment response up to this point.

Hasan et al. [28] cover emerging and existing biomarkers in this article that may be vital for early diagnosis, prognostic, and predictive indications of pancreatic cancer. CD73, a protein that attenuates tumor immunity, has been studied for its role in PC.

Chen et al. [29] found that patients with higher CD73 expression had poorer overall survival and disease-free survival. CD73 was also associated with reduced methylation and higher PD-L1 expression. This suggests that CD73 may be a biomarker for response to anti-PD-1/PD-L1 treatment in PC. Blockade of CD73 could be a promising therapeutic strategy for PC.

The results that we have seen in the literature indicate that different biomarkers are employed to identify PC. Numerous publications present systems for PC categorizations and early detections is done by using noncoding RNSs and the biomarker CA 19-9 as input. Th3dese systems are based on various machine learning and deep learning techniques. Urine proteome indicators are the main focus of this proposed system.

III. Methodology

PC is a devasting disease with a high risk rete, predicting at early is very crucial for patient.Accurate diagnosis is improve the chances to give perfect treament,recently ML techniquies have gained prominence in the health sector. By consider the publicly available dataset urinary biomarkers related to PC from Kaggle repository. To create any model frist step is the collection and preprocessing of data. After preprocessing step, data is splitted as training data and testing data(80% & 20%) .The parameters which influences the performance of the model are tuned by the hyperparameter tuning method gridsearchCV, these parameters are pass to the model as input data.GridsearchCV helps to fine-tune the parameters and optimizing its performance.A hybrid model is developed for classification of PC, with random forest and Ada Boost giving a promising approach to enhance the accuracy of diagnosis.

The created hybrid model demonstrates the major influence on early PC detection and, in the end, enhances patient outcomes in the PC battle. Fig. 1 shows the symbolic representation for the suggested pancreatic cancer classification system.

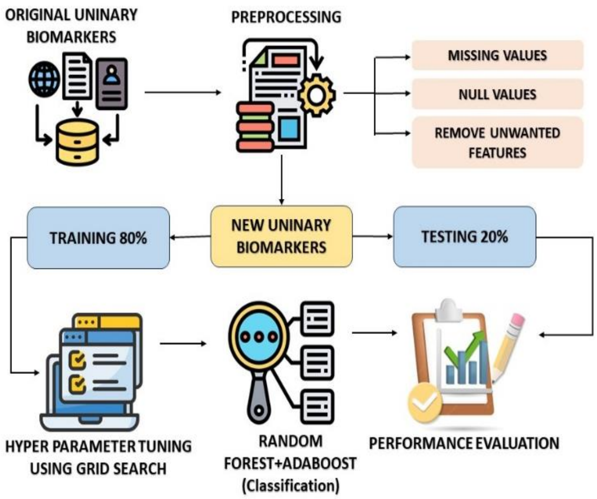


Fig. 1. Proposed pancreatic cancer classification system.

A. Dataset Description

The urinary biomarkers related to PC dataset are collected from the publicly available data repository Kaggle. Table I shows the urinary biomarkers dataset attributes and information. The data collection, supported by the biomarker panel, involved 590 urine samples gathered from various sources, including the University College London, the Barts Pancreas Tissue Bank, the University of Liverpool, Cambridge University Hospital, and the University of Belgrade, the Spanish National Cancer Research Centre. These samples comprised control samples 183, from individuals with 108 benign hepatobiliary conditions (including 119 from chronic pancreatitis patients), and 199 from individuals managed with pancreatic ductal adenocarcinoma (PDAC). The dataset has total 14 columns and 509 rows. The columns reflect numerous cancer risk factors, whereas the rows represent the study samples. The sixteen variables used in this data collection are as follows: the sample's ID. A cohort, which is a distinctive string used to identify each participant. Samples from cohort 1 have been utilized before. Samples from Cohort 2 have been added, and they are from the same above-mentioned sources. Sample_ Origin from where the sources are collected. Age, Sex, Diagnosis, Stage, Benign_Sample_Diagnosis, Blood plasma levels of the monoclonal antibody CA 19-9 are typically high in people with pancreatic cancer. There were just 350 subjects examined. Creatinine serves as an indicator of renal function in urine. Moreover, the presence of Lymphatic Vascular Endothelial Hyaluronan Receptor 1 (LYVE1) protein was identified in urine, indicating a potential significance in tumor metastasis. Additionally, the study quantified the urinary levels of proteins appended to pancreatic regeneration, specifically REG1A and REG1B. Furthermore, the urinary levels of Trefoil Factor 1 (TFF1), which might be related to the urinary tract, were assessed in a cohort of 306 patients.

Pancreatic cancer is categorized into stages such as IA, IB, IIA, IIB, III, and IV. However, in the dataset some input variables can be designated as irrelevant. Specifically, "sample id," automatically generated by Neural Designer, "Sample origin" signifies the source of patient samples, having no bearing on the ultimate diagnosis. "Stage" is a parameter exclusive to cancer patients. "Patient cohort" doesn't show any influence on the Sample Diagnosis, and "benign sample diagnosis" is an unrelated feature to the final sample diagnosis.

TABLE I. URINARY BIOMARKERS DATASET ATTRIBUTES AND INFORMATION

S. No	Feature	Specifies
1	Sample_id	Patient's id
2	Patient_cohort	To identify each participant
3	Sample_Origin	Where the patient samples came from
4	age	Patient's age, in years
5	sex	F-female M-male
6	diagnosis	Patient is having cancer
7	stage	Cancer stages IA, IB, IIA, IIIB, III, and IV
8	Benign_Sample_Diagnosis	Patient's not having cancer
9	plasma_CA19_9	Blood plasma levels of antibody CA_19-9
10	creatinine	Urine indicator of renal function
11	LYVE1	A protein in urine
12	REG1B	Pancreatic regeneration associated protein
13	TFF1	Urinary Trefoil Factor 1

The dataset consists of urine samples (590), classified into three variant groups of patients. These groups include healthy stage with 183 samples, and individuals with Stage of Benign and Stage of PDAC cases, 208 and 199 samples, respectively. Table II is a visual representation of this breakdown.

TABLE II. CHARACTERISTICS OF THE CLINICAL DATASET COMPRISING SAMPLES OF URINE FROM INDIVIDUALS WITH PANCREATIC CONDITIONS IN PROPOSED STUDY

Health Condition	Number of Sample	Gender	Range of age (median value)
Healthy stage	183	F (115) M (68)	26 – 89 (58) 30 – 87 (55)
Stage Benign	208	F (101) M (107)	26 – 82 (53) 29 – 82 (55)
Stage PDAC	199	F (83) M (116)	42 – 88 (68) 29 – 87 (67)

Fig. 2 shows the gender distribution of patients in different stages whereas Fig. 3 shows PC stage gender distribution on dataset.

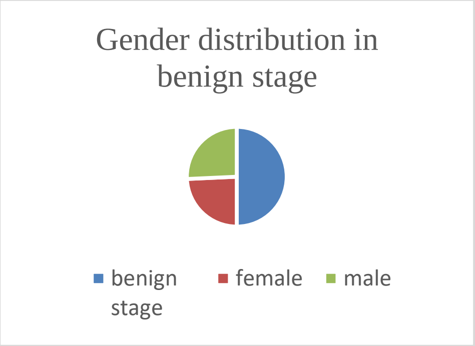


Fig. 2. Benign stage gender distribution on dataset.

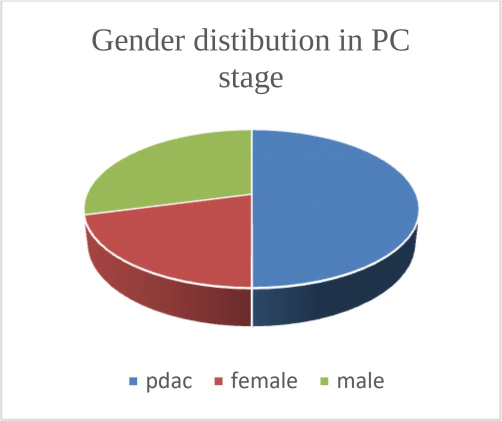


Fig. 3. PC stage gender distribution on dataset.

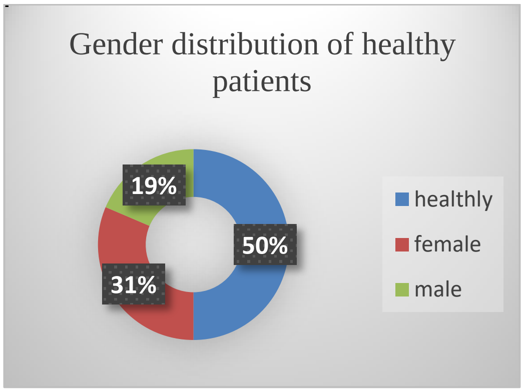


Fig. 4. Healthy patients gender distribution on dataset.

B. Preprocessing

Preprocessing is a crucial preliminary stage in data analysis. The situations where the dataset contains irrelevant, incomplete (missing), or inconsistent data, preprocessing becomes necessary. The preprocessing stage typically includes the following steps:

1) Data cleaning: This involves addressing missing values, which can be done by either omitting the entire data entry, replacing the missing value with a specific value, or employing strategies to manage null values. Inconsistencies in the dataset may also be resolved manually.

2) Data transformation: Converting data from one format to another format is called data transformation. When required, numerical data may undergo transformations for categorical variables, one-hot encoding is often employed to convert them into a suitable format for analysis.

In this study at data cleaning step the number of unwanted data, null data and missed data were cleaned. At transformation step one-hot encoding is used to covert char data into integer data.

Fig. 5 displays the graph that illustrates the correlation between each feature in the dataset and shows the degree of dependence between a variable to other variable used in the study. According to the color scale, the correlation relationship between the columns close to white is high, while the correlation relationship decreases gradually in the colors towards red.

C. Random Forest (RF)

RF is one of the best classification algorithms in Machine-learning. Instead of relying solely on output of a one decision tree, the random forest algorithm aggregates predictions from multiple decision trees then ultimately makes predictions based on the majority vote among these individual tree predictions. Fig. 6 shows random Forest classification. RF is a classification method that consists of a collection of decision trees, each built on different subsets of the dataset. By aggregating the results from these trees, often by taking their average, it enhances the predictive accuracy for the given dataset. But, the classification of random forest shows the low accuracy for the given dataset.

In RF classification Gini index is decide how nodes are distributed on a decision tree branch. It determines the branches is more likely to occur and find the class and probability of a node in each branch.

$$\text{Gini} \ = 1 - \sum_{i=1}^{j} f(i)^2 \qquad (1)$$

f is the class of frequency in the dataset.

c is the number of classes.

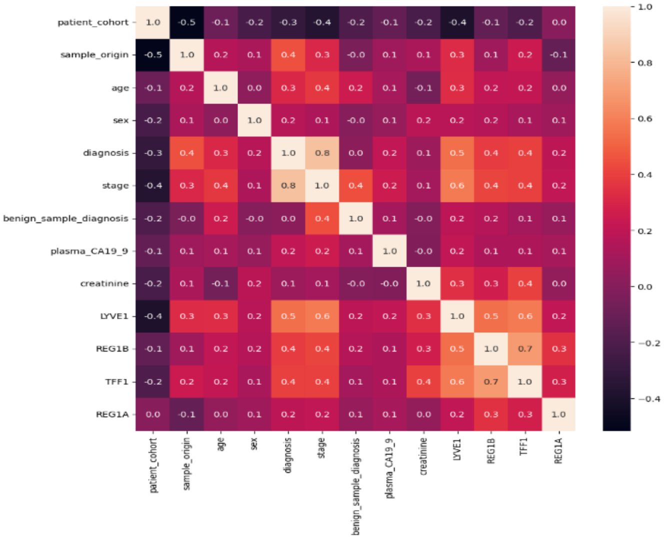


Fig. 5. Matrix of correlation.

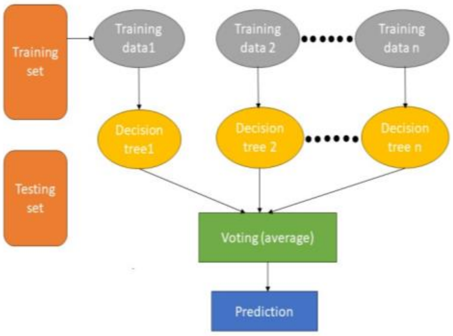


Fig. 6. Random forest classification.

There is another way to determine how nodes branch in a decision tree are determined is entropy. Entropy determines which branch the node should follow by analyzing the probability of a specific outcome. Its calculation involves the use of a logarithmic function, making it more mathematically sophisticated than the Gini index.

$$\text{Entropy } = \sum_{i=1}^{c} -p_i \log(p_i) \tag{2}$$

where, the p_i is the probability of randomly selecting a node in class i.

Information gain is a metric used to identify the most informative features in a dataset, and it is depended on the entropy value. It quantifies the difference in entropy before and after a data split and, in doing so, measures the impurity or disorder within class elements. Essentially, information gain helps determine which feature, when used to split the data, leads to the greatest reduction in uncertainty or entropy, making it a valuable criterion for feature selection in decision tree-based algorithms.

Information Gain= Entropy before splitting- Entropy after splitting

Consider the following pseudo code for RF classification.

Pseudo code of random forest classification.

Step 1: F features are select from the feature set randomly.

Step 2: In F each of x.

- Find the Gain.

$$\text{Gain } (s,x) = E(s) - E(s,x)$$

$$E(s) = \sum_{i=1}^{c} -f_i \log(f_i)$$

$$E(s,x) = \sum_{c \in X} P(c)E(c)$$

where E(s) two classes entropy, E(s,x) is the entropy of feature x..

- Choose the node with the highest information gain, denoted as y.
- Divide the node into sub-nodes.
- To build the tree, iterate through steps a, b, and c until the minimum required number of samples for splitting is reached.

Step 3: To create a forest consisting of n trees, replicate steps 1 and 2 n times.

The performance of RF model with all selected features is 83%. To improve the classification accuracy hyper parameter tuning is applied to evaluate the optimal parameters of the model. To find the best hyper parameters to build a single model which is suitable to improve the RF classification Grid search CV is one of the suitable hyper parameter techniques for the RF classification.

D. Hyperparameter Tuning

Hyperparameter tuning may be automated with the use of GridSearchCV, which also improves model performance and does away with manual trial-and-error. Fig. 7 shows grid search across two parameters. Grid Search combines each of the supplied hyperparameters and their values in a different way, calculates the performance of each combination, and then chooses the hyperparameters with the best value.

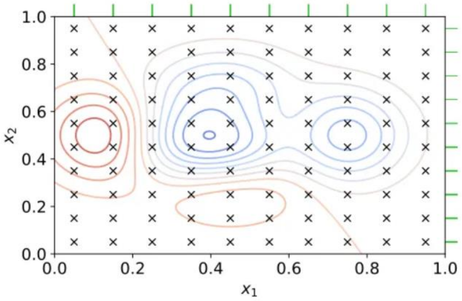


Fig. 7. Grid search across the two parameters.

Within the scikit-learn library's model_selection module, you can find a function known as GridSearchCV ().

Initiating the execution can be done by creating a GridSearchCV () object. In the scikit-learn model_selection class, there exists a function known as GridSearchCV (). The process begin can be begin by creating a GridSerachCV () object.

Clf=GridSearchCV (estimator, param_grid, cv, scoring).

The necessary four inputs are estimator, param_grid, cv, scoring.

These arguments are explained as follows:

1) estimator: A scikit-learn model

2) param_grid: A dictionary associating parameter names with lists of parameter values.

3) cv: A numeric value that specifies the number of folds in K-fold cross-validation.

4) scoring: Performance measure.

By using the GridSearchCV () prepare a best random model for predicting the pancreatic cancer efficiently. To obtain the better performance develop a hybrid model with the grid search cv random model with the help of Boosting algorithms. Ada Boost is one of the best boosting algorithms to improve the classification performance.

E. Boosting Algorithm

Ada Boost is one of the boosting algorithm in this one common estimator is used to learn for the decision trees in every split. By using with the splits, it builds a model with same weights in all levels if there are any high weights are assigned then those are called wrongly classified. In this context give the importance for high weights training with another model until reaches the minimum error. The argument is that while AdaBoost can be used independently of decision trees, it is frequently combined with them in reality because of the advantages that stumps have over other weak learners.

The mathematical summary of Ada Boost algorithm as follows:

1. Initializing Weights

   Dataset with samples of N, give each data point weight with $w_i = 1/N_s$.

   For each m=1 to M:

   Dataset sampled by the weight $w_i^{(m)}$ to attain training samples T_s

For each datapoint, AdaBoost initializes a weight of $1/N_s$. Once more, the weight of a datapoint indicates the likelihood that it will be chosen during sampling.

2. Training Weak Classifiers

   For all the training samples T_s , fit a classifier X_m

   X_m is a weak classifier then trained using the training dataset.

3. Here the AdaBoost algorithm is begins. The weights are updated as follows

$$\epsilon = \frac{\sum_{yi \neq K_m(x_i)} w_i^{(m)}}{\sum_{yi} w_i^{(m)}} \tag{3}$$

y_i is target variable ground truth value

w_i^m is the sample weight i at m^{th} iteration

the weight is updated at certain iteration m

$$\alpha_m = \frac{1}{2} \ln \frac{1-\epsilon}{\epsilon} \tag{4}$$

$$w_i^{(m+1)} = w_i^{(m)} e^{-\alpha_m y X_m(T_s)} \tag{5}$$

y represents the true values of the targent feature

$X_m(T_s)$ prediction made by the stump at iteration m

α_m predictive power of stump m

4. Making new overall Predictions

$$K(x) = \text{sign}[\sum_{m=1}^{M} \alpha_m X_m(x)] \tag{6}$$

The proposed hybrid classification model using a combination of a Random Forest classifier and an AdaBoost classifier developed. The Random Forest classifier is first tuned using gridsearchCV, and then the AdaBoost classifier is used to enhance the accuracy of RF. Finally, the working manner of the combined model is evaluated using metrics like accuracy, a classification report, and a confusion matrix.

IV. Results and Discussions

There are several researches are going on urinary biomarkers to determine whether they belonged to the pancreatic cancer or healthy patients (see Fig. 4). The publicly available dataset is used for the implementation of the proposed approach. The proposed system uses the hyperparameter tuning technique gridsearchCV to select the hyperparameters from the dataset, then perform the classification with RF and use AdaBoost to improve the classification accuracy. Training data make up 80% of the dataset, whereas testing data make up 20%. There are three subsections within the section: evaluation metrics, results, and comparison with other results of other approaches. This study employs classification algorithms and assesses their performance using metrics such as Accuracy, Recall, Precision, and the F-1 score.

Let TP (True Positive) represent the outcome in which the model correctly predicts the positive class, TN (True Negative) represents the outcome in which the model correctly predicts the negative class, FP (False Positive) represent the outcome in which the model incorrectly predicts the positive class, and FN represent the outcome in which the model incorrectly predicts the negative class. The above-mentioned performance measurements may thus be described as follows:

1) Accuracy: Accuracy, one of the most straightforward classification metrics, is computed as the ratio of correct predictions to all predictions made.

For calculation:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

2) Recall: True Positives represent predictions that correctly match the total number of positive cases, whether they were accurately predicted as positive or incorrectly predicted as negative (True Positives and False Negatives). The following is the calculation method for recall:

$$\text{Recall}= \frac{TP}{TP+FN}$$

3) Precision: Precision measures the proportion of positive predictions that were accurate. It can be expressed as the ratio of all correctly predicted positive instances (True Positives and False Positives) to the total number of positive predictions.

$$\text{Precision}= \frac{TP}{TP+FP} \times 100$$

4) F1-Score: Calculating the F1 Score involves taking the harmonic mean of Precision and Recall, giving equal weight to both metrics. The formula for computing the F1 score is as follows:

$$\text{F1- score}=2*\frac{precision*recall}{precision+recall}$$

5) ROC Curve: The classification accuracy also computed by the Receiver-Operating-Characteristic curve. It's constructed using True-Positive (TP) and False-Positive (FP) at different target levels. TP is determined by recall, while FP is established using fallout. This illustrates how a ROC curve plots the fallout and sensitivity of the classifier. Fig. 8 shows ROC analysis of the proposed system.

6) Confusion matrix: A confusion matrix is a common tool used in machine learning and statistics to evaluate the performance of a classification algorithm, particularly in the context of binary classification. It provides a clear and detailed breakdown of how a classifier's predictions align with the actual class labels in a dataset. A confusion matrix (see Fig. 9) is typically represented as a 2x2 matrix, and it contains four key metrics: TP, TN, FP, FN.

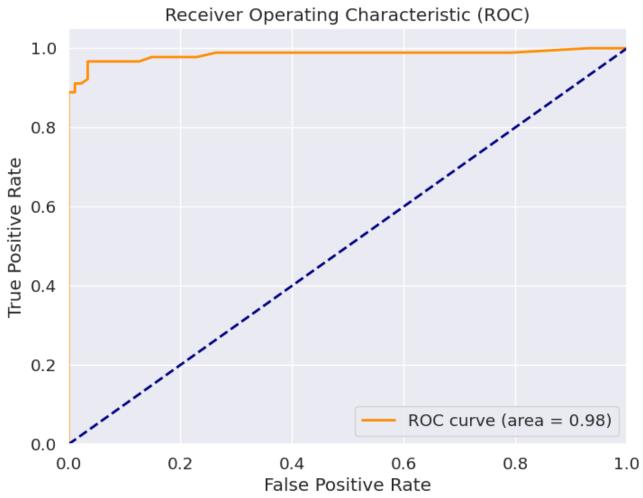


Fig. 8. ROC analysis of the proposed system.

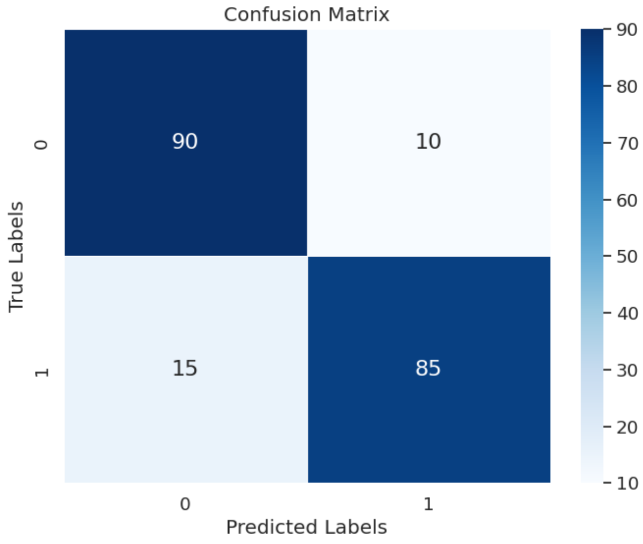


Fig. 9. Confusion matrix of proposed system.

Lately, machine learning techniques have seen substantial adoption in healthcare, particularly in the realm of cancer diagnosis. Utilizing hybrid ensembled classifiers, creatinine, LYVE1, REG1B, and TFF1 urine biomarkers have been effectively analyzed to detect pancreatic cancer. The proposed hybrid ensembled model surpasses other basic models with the maximum accuracy score of 97% in prior studies, as shown in Table III's assessment findings. Conventional ML models, such as LR, RF, and SVM, performed inadequately to correctly detect conditions related to pancreatic cancer. Therefore, hybrid ensembled classification model has been developed to predict pancreatic cancer based on urine biomarkers. As described above, the advantageous of gridsearchCV have the capability to ignore unnecessary feature and tune the feature which are very efficient to detect pancreatic cancer from the urine biomarkers.

7) Comparative analysis: The Table III offers an extensive comparison of the proposed model with established methods, conducting a thorough analysis of classification outcomes. Utilizing state-of-the-art classifiers, the proposed system is comprehensively evaluated in terms of accuracy, recall, precision, and the F1 score. Fig. 10 shows the comparative analysis of proposed system upon different classifiers based on urinary biomarkers.

TABLE III. A COMPARATIVE EVALUATION OF THE PROPOSED APPROACH WITH DIFFERENT CLASSIFIERS FOR PANCREATIC CANCER CLASSIFICATION USING URINARY BIOMARKERS

Classifier	Recall	Precision	F1-score	Accuracy
Random Forest	87	79	82	75
LR	76	64	89	74
KNN	64	65	-	64
GBC	77	73	-	72
1D CNN	100	90	95	93
LSTM-1D CNN	100	96	98	97
GridsearchCV+ RF+Ada Boost	**99.98**	**99.98**	**99.98**	**99.98**

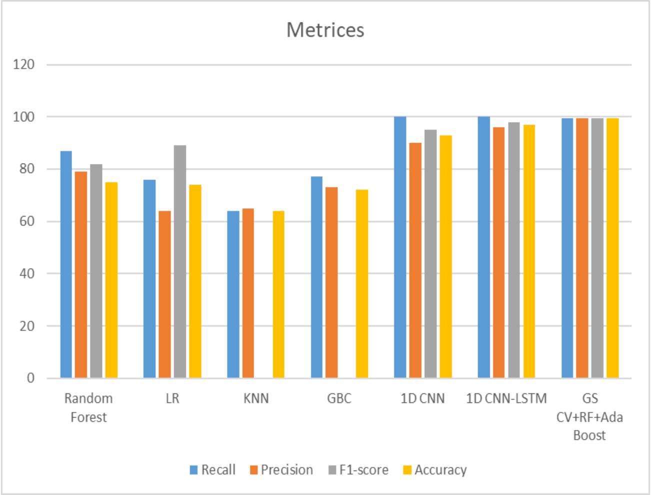


Fig. 10. Comparative analysis of proposed system upon different classifiers based on urinary biomarkers.

V. CONCLUSION

The scarcity of publicly available medical datasets is a significant challenge to the training of supervised learning models, particularly with regard to pancreatic cancer, as a result of which the only available training model performs poorly in terms of classification accuracy. A new ensemble strategy that uses urinary biomarkers gathered from Kaggle, all efficient features from dataset are tuning based on gridsearchCV then to classifies the pancreatic cancer and non-pancreatic cancer uses random forest, to improve performance of classification boosting algorithm Ada Boost was proposed. Metrics used to gauge the proposed model's performance Precision, F-1 score, Accuracy, and Recall all scored 99.98.

In future the machine learning models obtained optimum performances to detect pancreatic cancer at early stages automatically which are helpful for both clinicians and patients to save their life. This type of systems may assure prominent results in real time medical scenarios.

REFERENCES

[1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7–33.

[2] Chang YH, Margolin A, Madin O, et al. Deep learning-based nucleus classification in pancreas histological images. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2017, p. 672–675.

[3] Gupta, Anish, Apeksha Koul, and Yogesh Kumar. "Pancreatic cancer detection using machine and deep learning techniques." *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*. Vol. 2. IEEE, 2022.

[4] S. Fukushige and A. Horii, ‘‘Road to early detection of pancreatic cancer: Attempts to utilize epigenetic biomarkers,’’ Cancer Lett., vol. 342, no. 2, pp. 231–237, Jan. 2014.

[5] M. Kalubowilage et al., ‘‘Early detection of pancreatic cancers in liquid biopsies by ultrasensitive fluorescence nanobiosensors,’’ Nanomedicine, Nanotechnol., Biol. Med., vol. 14, no. 6, pp. 1823–1832, Aug. 2018.

[6] S. Boeck, D. P. Ankerst, and V. Heinemann, ‘‘The role of adjuvant chemotherapy for patients with resected pancreatic cancer: Systematic review of randomized controlled trials and meta-analysis,’’ Oncology, vol. 72, nos. 5–6, pp. 314–321, 2007.

[7] Malhotra A, Rachet B, Bonaventure A, Pereira SP, Woods LM. Can we screen for pancreatic cancer? Identifying a sub-population of patients at high risk of subsequent diagnosis using machine learning techniques applied to primary care data. PLoS One. 2021;16(6):e0251876.

[8] Karar ME, Alotaibi B, Alotaibi M. Intelligent medical IoT-enabled automated microscopic image diagnosis of acute blood cancers. 2022;22(6):2348

[9] Lee J, Lee HS, Park SB, Kim C, Kim K, Jung DE, Song SY: Identification of circulating serum miRNAs as novel biomarkers in pancreatic cancer using a penalized algorithm. Int J Mol Sci. 2021; 22:1007.

[10] Reddy, Santosh, and M. Chandrasekar. "PAD: A Pancreatic Cancer Detection based on Extracted Medical Data through Ensemble Methods in Machine Learning." *International Journal of Advanced Computer Science and Applications* 13.2 (2022).

[11] Lepowsky E, Ghaderinezhad F, Knowlton S, Tasoglu S. Paper-based assays for urine analysis Biomicrofuidics. 2017;11(5): 051501

[12] for urine analysis Biomicrofuidics. 2017;11(5): 051501. 17. Radon TP, Massat NJ, Jones R, Alrawashdeh W, Dumartin L, Ennis D, Dufy SW, Kocher HM, Pereira SP, Guarner L, et al. Identifcation of a three-bio- marker panel in urine for early detection of pancreatic adenocarcinoma. Clin Cancer Res. 2015;21(15):3512–21.

[13] Debernardi S, O'Brien H, Algahmdi AS, Malats N, Stewart GD, PlješaErcegovac M, Costello E, Greenhalf W, Saad A, Roberts R, et al. A combina- tion of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case–control study. PLoS Med. 2020;17(12): e1003489.

[14] Lee J, Lee HS, Park SB, Kim C, Kim K, Jung DE, Song SY: Identification of circulating serum miRNAs as novel biomarkers in pancreatic cancer using a penalized algorithm. Int J Mol Sci. 2021; 22:1007.

[15] Long NP, Jung KH, Anh NH, Yan HH, Nghi TD, Park S, Yoon SJ, Min JE, Kim HM, Lim JH et al: An integrative data mining and omics-based translational model for the identification and validation of oncogenic biomarkers of pancreatic cancer. Cancers. 2019; 11:155.

[16] Debernardi S, Massat NJ, Radon TP, Sangaralingam A, Banissi A, Ennis DP, Dowe T, Chelala C, Pereira SP, Kocher HM, et al. Non-invasive urinary miRNA biomarkers for early detection of pancreatic adenocarcinoma. Am J Cancer Res. 2015;5(11):3455–66

[17] Ko J, Bhagwat N, Yee SS, Ortiz N, Sahmoud A, Black T, Aiello NM, McKen- zie L, O'Hara M, Redlinger C, et al. Combining machine learning and nanofluidic technology to diagnose pancreatic cancer using exosomes. ACS Nano. 2017;11(11):11182–93

[18] Lee H-A, Chen K-W, Hsu C-Y: Prediction model for pancreatic cancer-A population-based study from NHIRD. Cancers. 2022; 14:882.

[19] Agarwal, Deepesh, et al. "Early detection of pancreatic cancers using liquid biopsies and hierarchical decision structure." *IEEE Journal of Translational Engineering in Health and Medicine* 10 (2022): 1-8.

[20] Thanya, T., and Wilfred Franklin S. "Novel computer aided diagnostic system using hybrid neural network for early detection of pancreatic cancer." *Automatika* 64.4 (2023): 816-827.

[21] Karar, Mohamed Esmail, Nawal El-Fishawy, and Marwa Radad. "Automated classification of urine biomarkers to diagnose pancreatic cancer using 1-D convolutional neural networks." *Journal of Biological Engineering* 17.1 (2023): 28.

[22] Blyuss O, Zaikin A, Cherepanova V, Munblit D, Kiseleva EM, Prytomanova OM, Dufy SW, Crnogorac -Jurcevic T. Development of PancRISK, a urine biomarker -based risk score for stratifed screening of pancreatic cancer patients. Br J Cancer. 2020;122(5):692–6.

[23] Jiao, Yan, et al. "ITGA3 serves as a diagnostic and prognostic biomarker for pancreatic cancer." *OncoTargets and therapy* 12 (2019): 4141.

[24] Liu, Yawen, et al. "Circulating lncRNA ABHD11-AS1 serves as a biomarker for early pancreatic cancer diagnosis." *Journal of Cancer* 10.16 (2019): 3746.

[25] Iwano, Tomohiko, et al. "High-performance collective biomarker from liquid biopsy for diagnosis of pancreatic cancer based on mass spectrometry and machine learning." *Journal of Cancer* 12.24 (2021): 7477.

[26] Takahashi, Kenji, et al. "Circulating extracellular vesicle-encapsulated HULC is a potential biomarker for human pancreatic cancer." *Cancer science* 111.1 (2020): 98-111.

[27] Brancaccio, Mariarita, et al. "Cell-free DNA methylation: the new frontiers of pancreatic cancer biomarkers' discovery." *Genes* 11.1 (2019): 14.

[28] Hasan, Syed, et al. "Advances in pancreatic cancer biomarkers." *Oncology reviews* 13.1 (2019).

[29] Chen, Qiangda, et al. "CD73 acts as a prognostic biomarker and promotes progression and immune escape in pancreatic cancer." *Journal of cellular and molecular medicine* 24.15 (2020): 8674-8686.

An Efficient Honeycomb Lung Segmentation Network Combining Multi-Paradigms Representation and Cascade Attention

Bingqian Yang, Xiufang Feng, Yunyun Dong*
School of Software, Taiyuan University of Technology, Taiyuan Shanxi 030024, China

***Abstract*—Honeycomb lung is a pulmonary manifestation that occurs in the terminal stage of various lung diseases, which greatly threatens patients. Due to the different locations and irregular shapes of lesions, the accurate segmentation of the honeycomb region is an essential and challenging problem. However, most deep learning methods struggle to effectively utilize both global and local information from lesion images, resulting in cannot to accurately segment the lesion. In addition, these methods often ignore some semantic information that is necessary for the segmentation of lesion location and shape in the decoding stage. To alleviate these challenges, in this paper, we propose a dual-branch encoder and cascaded decoder network (DECDNet) for segmenting honeycombs lesions. First, we design a dual-branch encoder consisting of ResNet34 and Swin-Transformer with different paradigm representations to extract local features and long-range dependencies respectively. Next, to further combine the different paradigm features, we develop the feature fusion module to obtain richer representation information. Finally, considering the problem of information loss during the decoder, a cascaded attention decoder is constructed to aggregate the multi-stage encoder information to get the final segmentation result. Experimental results demonstrate that our method outperforms other methods on the in-house honeycomb lung dataset. Notably, compared with the other nine universal methods, the proposed DECDNet obtains the highest IoU (86.34%), Dice (92.66%), Precision (93.21%), Recall (92.13%), F1-Score (92.66%), and achieves the lowest HD95 (7.33) and ASD (2.30). In particular, our method enables precisely segmenting lesions under different clinical scenarios as well. Our code and dataset are available at https://github.com/ybq17/DECDNet.**

***Keywords*—Honeycomb lung; attention; convolutional neural network; transformer; image segmentation**

I. INTRODUCTION

Honeycomb lung, also called interstitial pulmonary fibrosis, is a disease where the lung tissue is destroyed and fibrosed. It exhibits distinctive honeycomb-like features, that seriously threaten patient's life [1-2]. Based on published literature, the annual incidence of honeycomb lung is estimated to be between 0.9 and 13 cases per 100,000 individuals [3]. The survival time from initial diagnosis to death is short, ranging from three to five years, with a poor prognosis and high mortality rate [4-5]. Early diagnosis is vital for improving patient prognosis and prolonging their survival [6]. With the development of radiological technology, computed tomography (CT) has become the gold standard for diagnosing honeycomb lung [7]. Information such as the size and location of lesions in CT images can help doctors identify honeycomb regions and make subsequent treatment plans.

In clinical practice, the contours of lesions are usually delineated by physicians. However, the unique characteristics of honeycomb lung make the delineation of lesions a challenging task for physicians. Honeycomb lesions usually have the following properties: (1) the location of the lesion is not fixed and the contour is complex. (2) The lesions are blended with surrounding normal tissue, resulting in blurred boundaries. (3) The size and shape of lesions often vary from person to person. The above properties make it very time-consuming for doctors to contour lesions manually. In addition, due to the differences in doctor experience, the quality of delineation of diseased regions is inequality. To alleviate the burden on doctors and aid in diagnosis, many studies pay attention to computer-aided automatic diagnosis of lesions [8-12]. However, due to the inherent properties of these methods, it is challenging to achieve an accurate diagnosis of the lesion tissue.

With the development of deep learning, convolutional neural networks (CNN) have achieved significant success in the field of medical imaging due to their powerful feature extraction capabilities [13]. Many CNN methods have been applied to medical image segmentation, such as U-Net [14], V-Net [15], and ResUnet [16] which have obtained excellent results. The above methods all adopt an encoder-decoder structure, where the encoder is used to extract feature information at different scales of the image, and the decoder restores the image according to the encoder information. Since the superiority of this architecture, a few variants based on it have a dominant position in segmentation tasks [17-18]. Despite these models having achieved significant performance, their performance is still restricted due to their inherent receptive field [19]. To overcome the limitations, some works try to integrate attention mechanisms and expand the receptive field to improve segmentation accuracy [20-22]. Although the above works improve the segmentation performance of the network, they still suffer from capturing insufficient long-range dependencies.

Recently, transformer has achieved promising results in the field of natural language processing [23]. Due to its ability to capture long-distance dependencies, it has attracted the attention of researchers, and more and more works attempts to apply it to the field of computer vision. Dosovioskiy [24] was the first to introduce the transformer into the image recognition task; he divided an image into non-overlapping patches instead of tokens and fed them into transformer. In order to further improve the performance of image tasks, multi-scale transformer has emerged. For instance, Swin-Transformer [25] and PVT [26] used sliding windows and pyramid architectures respectively to reduce computational cost. Inspired by these studies, we apply the transformer to the honeycomb lung image segmentation task, but the results are not satisfactory due to the limitations of only using self-attention, which restricts the acquisition of local information.

CNN and transformer focus on two aspects of image information. On the one hand, due to the use of convolutional operations with inductive bias, CNN has locality and translation invariance [27]. This property allows CNN to preferably extract local information, but it also limits the receptive field, resulting in cannot to extract global features. Many solutions have been proposed to solve this problem, such as dilated convolution [28], large kernel convolution [29], and pyramid pooling [30]. However, these approaches can only alleviate this problem and cannot completely solve it. On the other hand, transformer utilizes the self-attention mechanism to capture long-range dependencies, but it neglects local information, resulting in the loss of detailed features.

According to the above analysis, we believe that CNN and transformer can be combined to compensate for their weaknesses and obtain higher performance. Several methods have tried to combine CNN and transformer to get local information and long-range dependencies to segment specific medical images, such as TransUNet [31], Tfcns [32], and HiFormer [33]. However, these architectures have some drawbacks that limit their ability to achieve better performance: 1) They cannot effectively combine local and global information from CNN and transformer respectively. 2) They cannot properly aggregate multi-stage information during the decoding stage. Considering these problems, we propose a novel network named DECDNet that combines different paradigms representation to segment the honeycomb region. Concretely, we first design two encoder branches, one is CNN to extract local information, and the other is transformer to get long-range dependencies. Second, we develop a feature fusion module to efficiently combine features from different branches. Lastly, to better aggregate multi-stage information, we construct an attention cascade decoder called ACD. Our contribution can be summarized as five-fold:

- We innovatively segment large-scale honeycomb lung CT images and plan to open this dataset. Our dataset will be available at https://github.com/ybq17/DECDNet.
- In order to extract the local information and global information of the image, a dual-branch encoder composed of ResNet34 and Swin-Transformer is designed to obtain the multi-paradigms representation of the image.
- Considering the problem of insufficient feature fusion, we develop a feature fusion module for efficiently combining information from different branches.
- To track the information loss in the decoding stage, a novel attention-based cascade decoder is constructed to aggregate multi-stage encoder information.
- Experimental results demonstrate that our proposed DECDNet surpasses other segmentation models as well as adaptable to different clinical scenarios.
- The rest of this paper is organized as follows. We first review related work in Section II and describe the overall architecture in Section III. In Section IV, we evaluate the performance of DECDNet on the honeycomb lung dataset. In Section V we discuss the experimental results. We conclude the paper in Section VI.

II. Related Works

A. Honeycomb CT Image Segmentation

CT imaging is a common technology used in clinical practice to diagnose honeycomb lung [34]. Different from pulmonary function tests (PFTs) and composite physiologic index (CPI) assessment in judging the patient's condition [35-36], CT imaging quantifies the degree of fibrosis in the lung and helps physicians diagnose the disease more intuitively [37]. Currently, many related studies focus on automatically segmenting honeycomb regions in CT images to quantify the degree of lesions.

For automatic segmentation methods of honeycomb lung can be divided into the following two categories: traditional computer-based CT analysis and deep learning-based models. The first category uses pulmonary fibrosis disease programs such as CALIPER to perform quantitative analysis of lesions. For instance, Jacob et al., [38] extracted characteristic manifestations in CT images, such as honeycomb lesions, reticular patterns, and ground-glass opacity. Then they used an automated procedure to quantify the severity. Nakagawa et al., [39] deployed a computer-assisted method to quantify fibrosis based on the estimated area of the lesions. However, such methods heavily rely on subjective experience that may lead to suboptimal results. Compared with traditional auxiliary approaches, deep learning-based methods achieve end-to-end automatic segmentation, not only reducing dependence on feature selection but also achieving promising results. Handa et al., [40] developed new deep-learning software that can automatically identify and quantify subdivided parenchymal honeycomb lesions. Su et al., [41] proposed a network called RDNet that accurately segmented the honeycomb regions, and the corresponding Dice index was 0.747. Considering more noise and lower contrast honeycomb images, Wei et al.,[42] introduced the popular transformer to quantify the lesions and achieved remarkable performance. Motivated by these studies, our work emphasizes the use of deep learning-based methods to reduce the role of feature selection and perform end-to-end lesions segmentation.

B. Convolutional Neural Network

With the development of deep learning, convolutional neural networks (CNN) have achieved remarkable success in a variety of computer vision tasks. In the segmentation task, Long et al., [43] proposed a fully convolutional neural network that achieves pixel-level classification. Inspired by the success of FCN, several approaches were introduced to improve segmentation performance, such as dilated convolution [28] and context modeling [30]. Later, an encoder-decoder network called UNet [14] was proposed by Ronneberger for medical image segmentation. With the popularity of UNet, a series of U-shaped architectures have been developed to better segment medical images, such as UNet++ [44], ResUnet [16], etc. In order to further improve segmentation performance, attention mechanisms were introduced into UNet by Ozan [21]. Attention mechanisms can selectively balance the importance of different spatial locations in the feature maps, allowing the network to focus on relevant objective regions. This approach has shown great potential in improving segmentation performance.

C. Vision Transformer

Transformer was designed originally for Natural Language Processing (NLP) and has achieved remarkable progress in the field [23]. Instead of using convolutional operations, transformer uses self-attention to capture long-range dependencies. Inspired by the success of transformer in the NLP field, Dosovitskiy et al., [24] proposed the vision transformer (ViT) model, which applied transformer to visual tasks for the first time. ViT split the image into patches, similar to words, and fed them into the transformer. Experimental results demonstrated that ViT surpasses many CNN models and has become the backbone for vision tasks. However, ViT requires a large amount of data and high computational complexity. To address this issue, several methods try to reduce the computational cost of ViT. Wang et al., [25] proposed the Pyramid Vision Transformer (PVT), which is inspired by the CNN pyramid structure and reduced the computational cost by using a hierarchical feature extraction strategy. Liu et al., [26] proposed the Swin Transformer, which only focuses on each local window to reduce the computational cost to linear. Recently, massive work has applied transformer to medical image segmentation tasks. Chen et al., [31] proposed TransUnet, which was the first to apply transformer to medical image segmentation. It used a multi-layer CNN-transformer encoder to extract both local and global information and a CNN decoder to restore the size of the output. Swin-UNet [45] and DS-TransUnet [46] used a pure transformer encoder-decoder architecture for 2D segmentation but did not achieve significant performance improvements. UNETR [47] adopted a transformer encoder to extract information and a CNN decoder to obtain 3D segmentation results.

III. Methodology

In this section, we describe our proposed method. Firstly, we illustrate a dual-branch encoder composed of ResNet34 and Swin-Transformer to extract complementary local information and long-distance dependencies of the image respectively. In the ResNet34 branch, the texture and structural information of lesions are fully utilized. In the Swin-Transformer branch, more attention is paid to the global features of lesions in images such as location and size. Then, we describe a feature fusion module that can fuse information from different branches at different scales. Finally, we present an ACD decoder that alleviates information loss during the decoding stage. As shown in Fig. 1, the input CT image size is 224×224×3, and the output result is a binary lesion segmentation result. To learn different paradigms representation, the input image is extracted by ResNet34 and multi-level Swin-Transformer respectively to extract multi-scale features of the image. Then, multi-scale features from different branches are fused by four feature fusion modules to obtain richer information. Lastly, to further reduce the information loss during the decoding stage, the ACD decoder receives fusion features to get the final segmentation result. More details are described in the following subsection.

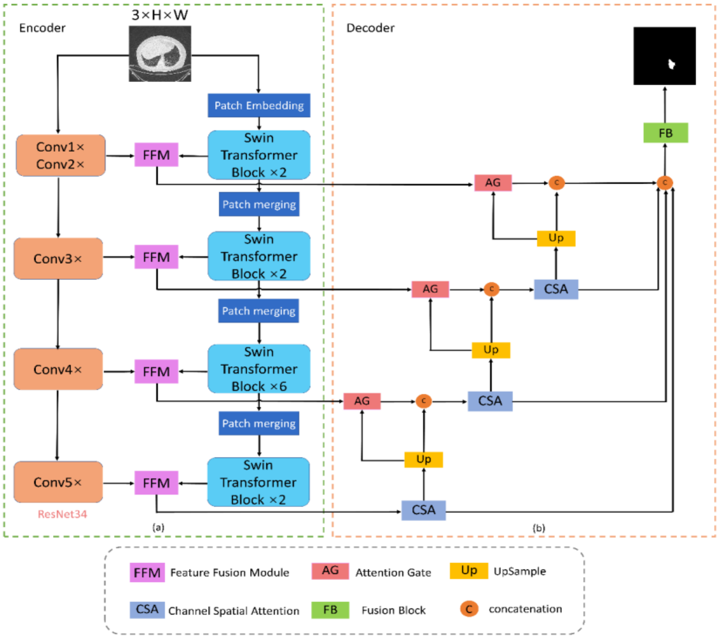


Fig. 1. The architecture of the proposed honeycomb lung segmentation network (DECDNet).

A. Encoder

Our proposed encoder consists of two multi-scale branches Resnet34, Swin-Transformer, and four feature fusion modules (FFM). Specifically, ResNet34 and Swin-Transformer respectively extract local and global information of images at different scales, such as texture, position, and structure. Here, ResNet34 and Swin-Transformer have four distinct layers each. To produce features with different scales, the patch merging operation will be performed before the feature map is inputted into the next Swin-Transformer layer. Then, the obtained multi-scale features through the above two parallel hierarchical branches contain different levels of semantic information. The FFM fuses each layer's information from different branches to enrich the features and feed them to the decoder.

1) CNN branch: As shown in Fig. 1(a), we use ResNet34 to construct the first encoder branch to obtain spatial detail and contextual information in the honeycomb lung images. In this branch, ResNet34 is divided into five blocks, noted as conv1, conv2, conv3, conv4, and conv5. When a feature map goes through a block, its width and height are reduced by a factor of 2. Taking a honeycomb lung CT image of size H×W×3 goes through conv1 and conv2 layers, which will yield a 3D feature c_1 of size H/4×W/4×C, where we set C to 48 to ensure feature consistency. Next, by passing through conv3, conv4, and conv5 layers, three features c_2, c_3, and c_4 are obtained, with sizes of H/8×W/8×2C, H/16×W/16×4C, and H/32×W/32×8C, respectively. These feature maps contain multi-scale semantic information and will be used as inputs for the feature fusion module.

2) Transformer branch: Recently, Alexely [24] proposed ViT, the first application of the transformer model to visual tasks. ViT has achieved outstanding progress in visual tasks due to its ability to capture long-range dependencies. It mainly consists of two parts: multi-head self-attention (MSA) and multi-layer perception (MLP). However, it has a major issue that its exponential computational complexity makes it difficult for downstream tasks such as image segmentation. To address this problem, Swin-Transformer [25] is developed that only focuses on local regions, greatly reducing the computational complexity. Inspired by the success of Swin-Transformer, we stack 12 Swin-Transformer blocks and patch operations to build the second encoder branch. In this branch, the focus is more on capturing the location and size information of the honeycomb lesions. Each Swin-Transformer is composed of two modules. The first module consists of W-MSA, LN, MLP, and residual connections. The second module introduces a sliding window mechanism to improve W-MSA and enhance information interaction. This process is defined as follows.

$$\hat{z}^l = W - MSA\left(LN(z^{l-1})\right) + z^{l-1},$$

$$z^l = MLP(LN(\hat{z}^l) + \hat{z}^l,$$

$$\hat{z}^{l+1} = SW - MSA(LN(z^l)) + z^l,$$

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \qquad (1)$$

The encoder branch is divided into four layers, as shown in Fig. 1(a). The first layer consists of two Swin-Transformer blocks and a patch embedding operation that splits the image into patches. The remaining three layers use patch merging to reduce the size of the image. And the number of Swin-Transformer blocks is set to 2, 6, and 2 for these layers, respectively. In this branch, the input image has the same dimensions as the CNN branch, which is H×W×3. The outputs of each layer are denoted as t_1, t_2, t_3, and t_4, with sizes of ((H/4, W/4), C), ((H/8, W/8), 2C), ((H/16, W/16), 4C) and ((H/32, W/32), 8C) respectively. They have the same pixel points as the outputs of the corresponding level CNN branch. We set the patch size to 4, so the feature dimension C is equal to 4×4×3=48.

3) Feature fusion module: In order to fuse different branches of features, we design the feature fusion module called FFM. It can efficiently and flexibly fuse features from CNN and transformer branches with different resolutions and different channels, and its structure is illustrated in Fig. 2. In this module, we first adjust the shape of the transformer features by reshaping operation. Then the CNN features and transformer features (called f and g) from each encoder branch are adjusted for dimensions by 1*1 convolution. Next, the f and g are merged by concatenation operation. The merged features are fed into the 1×1 convolution, and then the normalization operation and activation function are executed. Finally, the two encoder branch features are completely fused by a 1×1 convolution layer. In the encoder, we deploy four feature fusion modules to fuse the multi-scale features of CNN and Transformer under different branches. The fused feature balances multi-scale global information and local features, enriching the representation information.

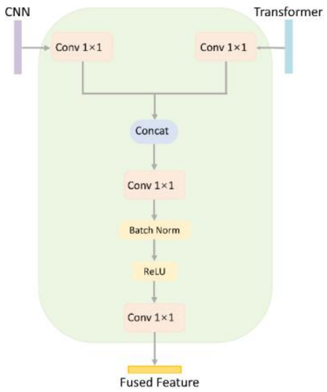


Fig. 2. The structure of the proposed Feature-Fusion Module (FFM).

B. Decoder

As illustrated in Fig. 1(b), we propose a novel decoder called ACD that can efficiently aggregate multi-stage information from the encoder. It consists of four components: Up, AG, CSA, and FB. Up is used for upsample operation, AG is used for cascaded feature fusion, CSA can refine features, and FB is used to fuse multi-level features. Specifically, we set up three CSA to enhance feature information at different scales and three AG corresponding to the output of FFB. In addition, AG is employed to integrate the fusion information from the FFB and the upsampled features from the lower level. Next, we use a concatenation operation to merge the features of AG and the lower layer. Then, CSA is executed to refine the merged features. Finally, we use FB to integrate the output of different layers to obtain the final prediction. More details are described in the following subsection.

1) Up: Up is used to restore the current feature size to match the dimension of the upper layer's feature. Each Up consists of a ReLU activation function, batch normalization operation, a 3*3 convolution operation, and a linear upsampling. Upsample () with an upsampling factor of 2. The Up operation can be formulated as follows:

$$Up(x) = (ReLu(BN(Conv(Upsample(x))))) \quad (2)$$

2) Fusion block: To enhance feature representation, we design the Fusion Block (FB) to efficiently utilize multi-level features. As shown in Fig. 3(a), FB consists of two parts: residual connection and Convblock. The Convblock consists of Convolution (Conv), Batch Normalization (BN), and Rectified Linear Unit (ReLu). On the one hand, the convolutional block assists the network in enhancing features. On the other hand, the residual connection avoids gradient explosion. Then, the results of the two parts are added together to obtain the final prediction result.

3) Channel spatial attention: The Channel Spatial Attention (CSA) module can refine the feature map, as shown in Fig. 3(b). It consists of spatial attention, channel attention, and a 1×1 convolution. When the input features enter this module, parallel channel attention and spatial attention are used to enhance the spatial and channel features of the image, respectively. Then, these features are fused through concatenation and the dimension is reduced through convolution. The fusion feature retains important spatial and channel information, which assists the decoder in better-recovering information.

4) Attention gate: Motivated by the success of the attention mechanism, we introduce the attention gate (AG), which can combine multi-stage features, extract areas of interest, and ignore irrelevant parts using spatial attention. As shown in Fig. 3(c), each AG consists of two 1x1 convolutions to change dimensions, two batch normalizations, and two activation functions: ReLu and Sigmoid. Specifically, the FFB features denoted f is first added point-wise with the features from the lower-level denoted d. Then, convolution and activation operations are performed to obtain the spatial attention map. Finally, the attention map is element-wise multiplied by f using the Hadamard product.

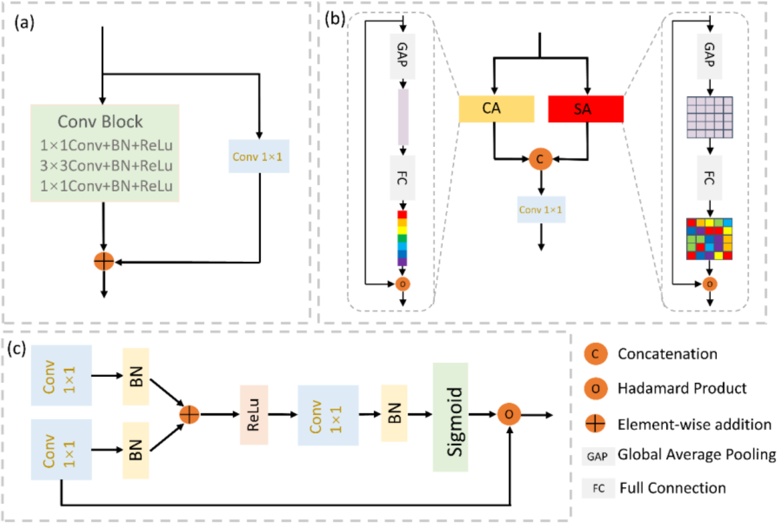


Fig. 3. Structure of (a) FB, (b) CSA, and (c) AG.

IV. Experimental Results

A. Datasets

We use CT images of honeycomb lungs provided by Shanxi Provincial People's Hospital for the experiment. A total of 7170 honeycomb images with size 512 × 512 pixels were collected from chest CT scans of 121 patients as our dataset. These scans were conducted using specific scan parameters, including an X-ray voltage of approximately 120 kVp, a current of 200-500 mA, and a gantry rotation speed of 0.5 seconds per rotation. And the slice images were acquired with a uniform slice thickness of 5 mm. The dataset includes images with lesion sizes ranging from small focal areas of a few millimeters to extensive lesions spanning large lung areas. Lesion appearances also vary, from early-stage subtle honeycombing with fine reticular patterns to advanced-stage prominent cystic changes as illustrated in Fig. 4. All of these images are labeled by experienced radiologists. After annotation, we resize the images to 224×224 and apply normalization to accelerate model convergence. We divide the dataset into training/validation/testing sets at a ratio of 6:2:2, with 4302 images used for training and 1434 images used for validation and testing.

B. Evaluation Metrics

We adopt five common metrics to evaluate segmentation performance: Dice coefficient (Dice), Jaccard index (IoU), Precision, Recall, and F1-score. These metrics are calculated using true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN):

$$Dice = \frac{2TP}{2TP+FP+FN} \tag{3}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

$$F1-score = \frac{2\times Recall\times Precision}{Recall+Precision} \tag{7}$$

In addition, Hausdorff distance (HD) 95% and average surface distance (ASD) are used to measure the performance of the model in segmenting the boundaries of honeycomb regions.

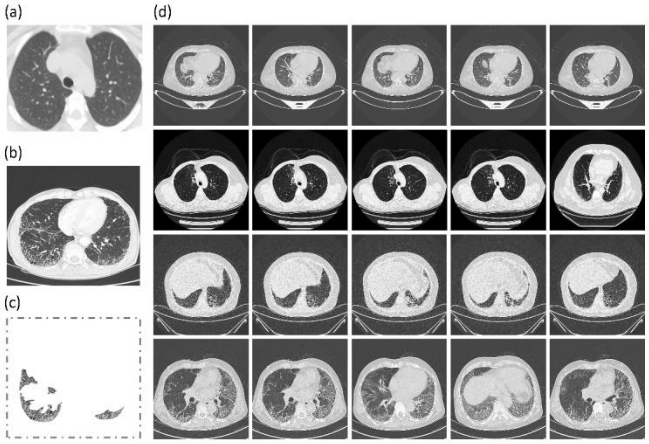


Fig. 4. (a) A normal lung image, (b) A honeycomb lung image, (c) Honeycomb lung lesions, and (d) Some honeycomb lung images in the datasets.

C. Implementation Details

Our architecture is implemented using PyTorch and trained on an RTX Nvidia 3090 GPU. And we use the SGD as the optimizer with a momentum of 0.9 and weight decay of 0.0001. The learning rate and batch size are set to 0.0001 and 24, respectively. Further, aiming to improve the generalization and robustness of the model, we use data augmentation technology such as flipping and rotation to diversify the data.

The loss function is composed of commonly used cross-entropy loss and dice loss, defined as follows:

$$= \alpha l_1 + (1-\alpha) l_2 \tag{8}$$

where, l_1 denotes the cross-entropy loss and l_2 denotes the dice loss. The weight ratio α is a balancing factor, set to 0.4 based on experimental testing.

D. Evaluation Results

To demonstrate the effectiveness of our proposed DECDNet, we conducted comparative experiments with nine methods. All methods are evaluated both quantitatively and qualitatively. These competing methods include the following: U-Net [12], Att-UNet [19], UNet++ [42], FPN [28], DABNet [46], ViT [22], CGNet [47], TransUNet [29], SwinUNet [43]. The above methods cover three types: traditional CNN architecture, pure transformer architecture, and the combined architecture of CNN and transformer. And we evaluated our method on seven evaluation metrics mentioned earlier. The smaller the ASD and HD95 value is, the better the performance is; the larger the other index values are, the better the performance is. Table I shows the qualitative evaluation results of all methods on the honeycomb lung dataset. The experimental results demonstrate that our proposed network has higher IoU, Dice, Precision, Recall, and F1-score, as well as smaller HD95 and ASD. Compared with the second-ranked SwinUNet, our method increases IoU and Dice by 2.13% and 1.23%, respectively, reaching 86.34% and 92.66% Precision, Recall, and F1-Score also improved by 1.68%, 0.81%, and 1.24%, respectively. HD95 and ASD decreased by 2.08 and 0.51 compared to Swin-UNet, reaching 7.33 and 2.30, respectively. These results indicate that our proposed DECDNet can accurately segment the honeycomb lesions. Moreover, our method outperforms traditional CNN methods, ViT, and TransUnet in terms of FLOPs or Params, not only achieving the best segmentation results but also balancing the accuracy and computational complexity.

Meanwhile, to quantify segmentation performance, we conducted visual comparisons of the segmentation with different methods on the honeycomb lung dataset. As shown in Fig. 5, we visually compared the segmentation results of UNet, FPN, TransUNet, SwinUNet, and our model with the ground truth. Within these results, the red box draws attention to significant differences compared with the mask. Obviously, the transformer method is more accurate than the traditional CNN method for segmenting honeycomb lung lesions However, due to the transformer ignores the local features of the image, the segmentation of the lesion boundary is smoother compared with the mask. In contrast, our proposed architecture efficiently utilizes different paradigms representation and alleviates the information loss during the decoding stage. The boundary delineation and region segmentation of the honeycomb lung are more similar to the ground truth. In the above quantitative and qualitative analyses, DECDNet achieves the best performance on the honeycomb lung dataset, demonstrating that our proposed method can accurately segment the contours and regions of honeycomb lung.

Additionally, to evaluate the adaptability of this model in clinical practice, we performed experiments to evaluate our method under various conditions considering the patient positioning, lesion size, and the presence of adjacent organs that commonly occur in clinical applications. We conducted the visual analysis of CT images of different lesion sizes, different slices, and different angles. As shown in Fig. 6 and Fig. 7, regardless of the size, axis, or slice of the lesion, DECDNet can accurately segment and outline the lesions. Table II shows the quantitative results for Dice and Hd95 in various conditions, further proving the adaptability of our method for lesion area segmentation in different clinical scenarios.

TABLE I. COMPARING OUR METHOD WITH OTHER METHODS ON THE HONEYCOMB LUNG DATASET

Methods	IoU (%)	Dice (%)	Precision (%)	Recall (%)	F1-Score (%)	HD	ASD	FLOPs(G)	Params(M)
U-Net [14]	76.82	86.89	87.58	86.20	86.38	15.05	5.27	37.03	31.04
Att-UNet [21]	78.12	87.72	87.21	88.23	87.71	14.20	4.54	64.19	57.16
U-Net++ [44]	79.67	88.68	87.71	89.68	88.68	13.69	4.13	103.36	47.19
FPN [30]	81.65	89.90	89.89	89.46	89.67	12.11	3.86	38.49	36.77
DABNet [48]	82.08	90.16	90.90	89.32	90.10	11.58	3.52	1.01	0.75
ViT [24]	80.32	89.08	89.48	88.57	89.02	13.15	4.78	17.58	85.80
CGNet [49]	82.72	90.54	91.01	89.38	90.18	11.27	3.39	0.68	0.49
TransUNet [31]	83.53	91.03	91.55	90.49	91.02	10.36	2.97	29.35	105.28
SwinUNet [45]	84.21	91.43	91.53	91.32	91.42	9.41	2.81	6.16	27.17
Ours	86.34	92.66	93.21	92.13	92.66	7.33	2.30	17.62	90.37

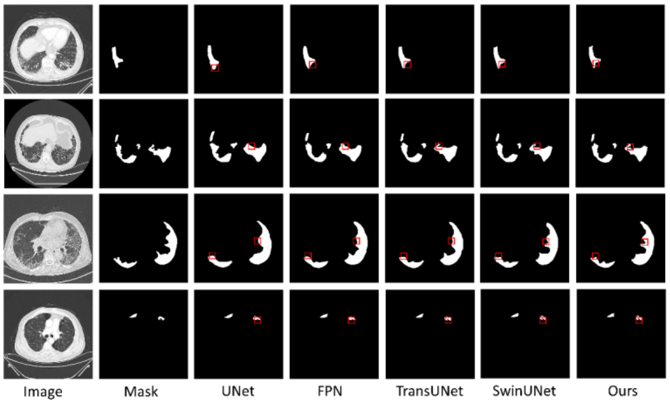
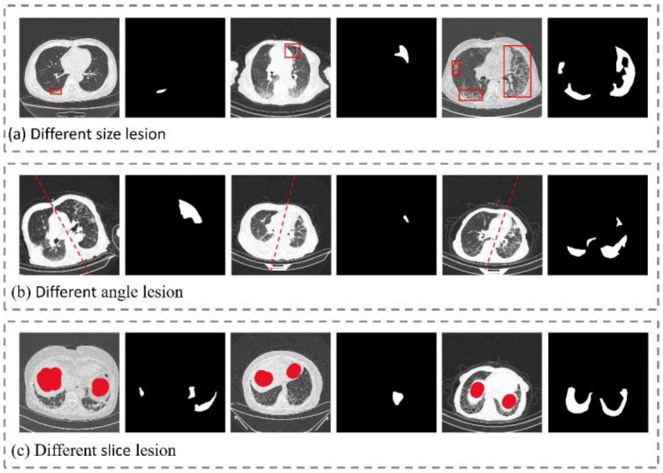


Fig. 5. The qualitative comparison of the honeycomb lung dataset.


(a) Different size lesion

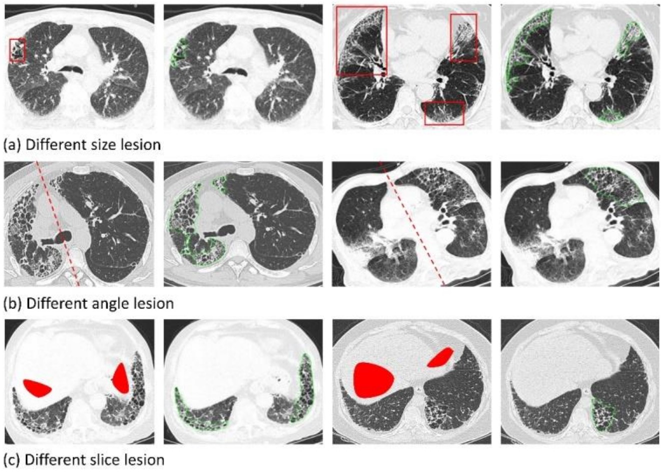
(b) Different angle lesion

(c) Different slice lesion

Fig. 6. DECDNet segmentation lesions visualization under different conditions (a) represents the lesion size (b) represents scan angle (c) represents the slice lesion.

TABLE II. QUANTITATIVE ANALYSIS UNDER DIFFERENT CONDITIONS

Condition	Dice	HD
Different size	91.24	7.61
Different angle	92.83	7.15
Different slice	92.70	7.26


(a) Different size lesion

(b) Different angle lesion

(c) Different slice lesion

Fig. 7. DECDNet segmentation lesion boundaries visualization under different conditions (a) represents the lesion size (b) represents scan angle (c) represents the slice lesion.

E. Ablation Studies

1) Evaluation of encoder: Our encoder is composed of CNN and transformer with different paradigms. Since the transformer can capture long-range dependencies, it tends to ignore local feature information. To address this issue, we introduced CNN as an auxiliary branch to compensate for the loss of spatial detail information. To further verify the importance of the CNN branch, we employed variants of pure transformer and CNN-combined transformer as two encoder structures. As shown in Fig. 8, the encoder that combines CNN with the transformer outperforms the pure transformer encoder. The results presented in Table III suggest that the integration of the CNN branch also enhances boundary segmentation performance. Thus, the introduction of CNN as an auxiliary branch is necessary to improve segmentation performance.

TABLE III. ABLATION EXPERIMENTS ON BOUNDARY EVALUATION FOR CNN BRANCH

Methods	HD	ASD
No-CNN	8.40	3.01
Ours	7.33	2.30

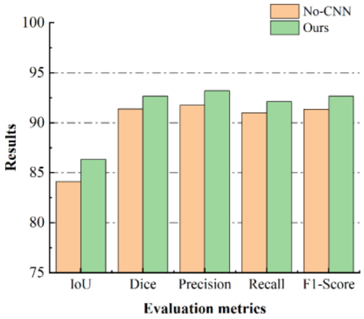


Fig. 8. Ablation study on the influence of CNN branch.

2) Evaluation of FFM: To verify the effectiveness of FFM in segmentation performance, we removed FFM and used Conv2d to adjust the feature dimensions, then adopted the add operation to fuse feature. In this way, we denote the variable as Conv2d+add to verify the importance of FFM. The segmentation performance shown in Fig. 9 reveals the effect of FFM in feature fusion. Specifically, the model with FFM achieved higher IoU, Dice, Precision, Recall, and F1-score compared to using only the add operation. In addition, the employment of FFM can improve boundary segmentation performance. As shown in Table IV, our method achieved lower HD95 and ASD. These results demonstrate that simply using add cannot fully utilize the features from CNN and transformer, while FFM can better help the network utilize different branches feature.

3) Evaluation of decoder: The decoder is an important factor that affects segmentation performance, and its main role is to restore feature maps to obtain the final prediction. The classic decoder does not use attention mechanisms but restores image resolution by fusing multi-stage skip connections and upsampling features to achieve the final prediction. While, Atten-UNet uses attention in skip connections during the decoding stage to focus on the lesions, denoted as the attention decoder. Different from the above decoder, we proposed a new attention-based cascaded decoder called ACD to efficiently combine multi-stage features from the encoder, enabling the network to better focus on the honeycomb regions. The results shown in Fig. 10 demonstrate that compared to the classic decoder and attention decoder, the architecture using ACD achieves improvements in IoU, Dice, Precision, Recall, and F1-score. Our decoder also obtains the smallest HD95 and ASD, as shown in Table V. The presented results provide evidence that our decoder can better focus on the lesions and efficiently integrate multi-stage features to get segmentation results.

4) Evaluation of combination FFM and cascade: To verify the effectiveness of the FFM and Cascade decoder combination for model performance, we conducted four sets of experiments to explore the influence of each component on honeycomb lung segmentation. Firstly, setting a model that only uses pointwise additive fusion and classical encoder for segmentation as the baseline model. To ensure fairness, all experiments adopt the same environment settings.

TABLE IV. ABLATION EXPERIMENTS ON BOUNDARY EVALUATION FOR FFM

Methods	HD	ASD
Conv2d+add	8.27	3.09
Ours	7.33	2.30

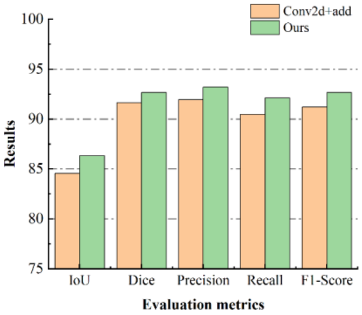


Fig. 9. Ablation study on the influence of FFM.

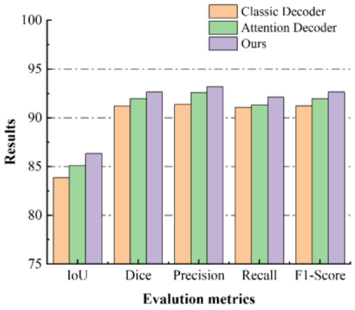


Fig. 10. Ablation study on the influence of different decoders.

TABLE V. ABLATION EXPERIMENTS ON BOUNDARY EVALUATION FOR DIFFERENT DECODERS

Methods	HD	ASD
Classic decoder	11.30	3.38
Attention decoder	8.19	2.88
Ours	7.33	2.30

As shown in Fig. 11, the combination of FFM and Cascade decoder achieved the best segmentation performance of honeycomb lesions. For boundary outline, the combination method also obtains the lowest HD and ASD as illustrated in Table VI, which is more similar to the groundtruth. Therefore, FFM and Cascade are necessary to improve the accuracy of segmentation.

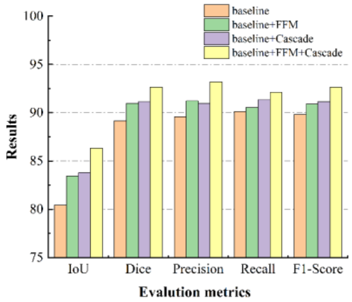


Fig. 11. Ablation study on the influence of the combination FFM and Cascade.

TABLE VI. ABLATION EXPERIMENTS ON BOUNDARY EVALUATION FOR THE COMBINATION FFM AND CASCADE

Methods	HD	ASD
baseline	12.78	4.05
baseline+FFM	10.60	3.29
baseline+Cascade	9.82	3.11
baseline+FFM+Cascade	7.33	2.30

V. DISCUSSION

Honeycomb lung is a terminal manifestation of lung disease, which greatly threatens patients. In clinical applications, the segmentation of lesions is essential. It aids in evaluating lesions, identifying the distribution of lesions, and assisting doctors in making accurate diagnoses. Moreover, segmenting honeycomb lung is a challenging task due to their ambiguous and irregular characteristics. Therefore, it is crucial to design a network that achieves higher segmentation accuracy for the precise localization of honeycomb lung lesions. Our proposed DECDNet architecture integrates global and local information from different paradigms to alleviate the limitations of CNN and transformer. In addition, the specially designed ACD decoder can effectively recover image information from the encoder. We conducted experiments on our method and nine universal segmentation algorithms, and our method achieved the highest IoU (86.34%), Dice (91.87%), Recall (92.13%), Precision (93.21%), F1-score (92.66%), and the smallest HD95 (7.33) and ASD (2.30). To quantify the results, we show the visual segmentation results of different methods, as shown in Fig. 5. Compared with other methods, our method can not only focus on the major lesions but also pay more attention to the boundaries of the honeycomb lung. Next, we visualized the segmentation performance in different clinical situations, as shown in Fig. 6 and Fig. 7, indicating that the proposed model is adaptable to diverse clinical scenarios. Additionally, to verify the effectiveness of each part of our architecture, we conducted three groups of ablation experiments to explore the effects of the dual-branch encoder, FFM, and ACD decoder. The results show that all three parts are effective and can improve the segmentation performance of the network. Therefore, our proposed method can precisely segment the honeycomb lung lesions, alleviate the burden on doctors, and assist in diagnosis.

Although our method has shown outstanding performance on the honeycomb lung dataset, it still has some defects that need to be addressed. On one hand, our model did not achieve promising performance in cases where the lesion size is small and the background is complex. On the other hand, as we have only tested on data from a single center, the generalizability of our model remains to be considered. In the future, we will expand and improve our method by adjusting multi-scale inputs and collecting data from multiple centers to address these issues.

VI. CONCLUSION

In this paper, we propose a novel network called DECDNet for the segmentation of honeycomb lung CT images. Specifically, we first design a dual-branch encoder to efficiently capture global and local information from different paradigms. Next, the feature fusion module is developed to fuse CNN and transformer features. Finally, we develop an attention-based cascade decoder to aggregate multi-stage encoder information. Our method demonstrated its effectiveness in extensive experiments through the effective extraction, fusion, and restoration of local information (such as the texture and structure of lesions) and global information (such as location and size). And our model achieves state-of-the-art performance on the honeycomb lung dataset. In addition, our model also accurately segments lesions under various conditions, making it a valuable method for assisting doctors in locating and tracking lesions, as well as making diagnoses. In the future, we will focus on further automatic diagnosis of honeycomb lung, such as by adopting multi-scale inputs to avoid noise and collecting multi-center data to enhance the model's generalizability.

ACKNOWLEDGMENT

We thank all the anonymous reviewers for their valuable comments and constructive suggestions. This work is supported by the National Natural Science Foundation of China (No.62306206); the Key Research and Development Program of Shanxi Province (No.202102020101007).

REFERENCES

[1] M. Hosseini and M. Salvatore, “Is pulmonary fibrosis a precancerous disease?” European Journal of Radiology, p. 110723, 2023.

[2] G Raghu, M Remy-Jardin, L Richeldi, C. C. Thomson, Y. Inou, and T. Johkoh, “Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline,” American Journal of Respiratory and Critical Care Medicine, vol. 205, no. 9, pp. e18–e47, 2022.

[3] T. M. Maher, E. Bendstrup, L. Dron, J. Langley, G. Smith, and J. M. Khalid, “Global incidence and prevalence of idiopathic pulmonary fibrosis,” Respiratory research, vol. 22, no. 1, pp. 1–10, 2021.

[4] Q. Zheng, I. A. Cox, J. A. Campbell, Q. Xia, P. Otahal, and Bde. Graaff, “Mortality and survival in idiopathic pulmonary fibrosis: a systematic review and meta-analysis,” ERJ Open Research, vol. 8, no. 1, 2022.

[5] J. Ji, S. Zheng, Y. Liu, T. Xie, X. Zhu, and Y. Ni, “Increased expression of OPN contributes to idiopathic pulmonary fibrosis and indicates a poor prognosis,” Journal of Translational Medicine, vol. 21, no. 1, p. 640, 2023.

[6] M. B. Herberts, T. T. Teague, V. Thao, L. R. Sangaralingham, H. J. Henk, and K. T. Hovde, “Idiopathic pulmonary fibrosis in the United States: time to diagnosis and treatment,” BMC Pulmonary Medicine, vol. 23, no. 1, p. 281, 2023.

[7] S. Hobbs, J. H. Chung, J. Leb, K. Kaproth-Joslin, and D. A. Lynch, “Practical imaging interpretation in patients suspected of having idiopathic pulmonary fibrosis: official recommendations from the Radiology Working Group of the Pulmonary Fibrosis Foundation,” Radiology: Cardiothoracic Imaging, vol. 3, no. 1, p. e200279, 2021.

[8] P. Mathur, A. S. Raghuvanshi, A. Kumari, and A. Chandra, “Computer-Aided Diagnosis System for Brain Tumor Classification and Segmentation,” in 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), 2023, pp. 547–552.

[9] Y. Xia, H. Yun, and Y. Liu, “MFEFNet: Multi-scale feature enhancement and Fusion Network for polyp segmentation,” Computers in Biology and Medicine, vol. 157, p. 106735, 2023

[10] X. He, G. Qi, Z. Zhu, Y. Li, B. Cong, and L. Bai, “Medical image segmentation method based on multi-feature interaction and fusion over cloud computing,” Simulation Modelling Practice and Theory, vol. 126, p. 102769, 2023.

[11] Aamir M, Rahman Z, Dayo Z A, Abor W A, Uddin M, and Khan , “A deep learning approach for brain tumor classification using MRI

images," Computers and Electrical Engineering, vol. 101, p. 108105, 2022.

[12] Aamir M, Rahman Z, Abro W A, Bhatti U A, Dayo Z A, and Ishfaq M, "Brain tumor classification utilizing deep features derived from high-quality regions in MRI images," Biomedical Signal Processing and Control, vol. 85, p. 104988, 2023.

[13] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle, "Deep neural networks for medical image segmentation," Journal of Healthcare Engineering, vol. 2022, 2022.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 2015, pp. 234–241.

[15] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 fourth international conference on 3D vision (3DV), 2016, pp. 565–571.

[16] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-unet for high-quality retina vessel segmentation," in 2018 9th international conference on information technology in medicine and education (ITME), 2018, pp. 327–331.

[17] H. Sahli, A. Ben Slama, and S. Labidi, "U-Net: A valuable encoder-decoder architecture for liver tumors segmentation in CT images," Journal of X-ray science and technology, vol. 30, no. 1, pp. 45–56, 2022.

[18] S. Yalçın and H. Vural, "Brain stroke classification and segmentation using encoder-decoder based deep convolutional neural networks," Computers in Biology and Medicine, vol. 149, p. 105941, 2022.

[19] Y. Xie, J. Zhang, C. Shen, and Y. Xia, "Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24, 2021, pp. 171–180.

[20] J. Zhang, W. Pan, B. Wang, Q. Chen, and Y. Cheng, "Multi-scale aggregation networks with flexible receptive fields for melanoma segmentation," Biomedical Signal Processing and Control, vol. 78, p. 103950, 2022.

[21] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, and K. Misawa, "Attention u-net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

[22] D. P. Fan, G. P. Ji, T. Zhou, G. Chen, H. Fu, and J. Shen, "Pranet: Parallel reverse attention network for polyp segmentation," in International conference on medical image computing and computer-assisted intervention, 2020, pp. 263–273.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, and A. N. Gome, "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.

[24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, and T. Unterthiner, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.

[25] Z Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, and Z. Zhan, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 10012–10022.

[26] W. Wang, E. Xie, X. Li, D. Fan, K. Song, and D. Liang, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 568–578.

[27] A. Mumuni and F. Mumuni, "CNN architectures for geometric transformation-invariant feature representation in computer vision: a review," SN Computer Science, vol. 2, pp. 1–23, 2021.

[28] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv preprint arXiv:1511.07122, 2015.

[29] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters--improve semantic segmentation by global convolutional network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4353–4361.

[30] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.

[31] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, and Y. Wang, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.

[32] Z. Li, D. Li, C. Xu, W. Wang, Q. Hong and Q. L, "Tfcns: A cnn-transformer hybrid network for medical image segmentation," in International Conference on Artificial Neural Networks, 2022, pp. 781–792.

[33] M. Heidari, A. Kazerouni, M. Soltany, R. Azad, E. K. Aghdam, and J. Cohen-Adad, "Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 6202–6212.

[34] P. Li, J. Zhang, W. Yua, and S. Yu, "Idiopathic interstitial pneumonia," In: Radiology of Infectious and Inflammatory Diseases-Volume 3: Heart and Chest, 2023, pp. 309–323.

[35] R. Gupta, J. S. Kim, and R. P. Baughman, "An expert overview of pulmonary fibrosis in sarcoidosis," Expert Review of Respiratory Medicine, vol. 17, no. 2, pp. 119–130, 2023.

[36] Q. Wang, Z. Xie, N. Wan, L. Yang, Z. Jin, and F. Jin, "Potential biomarkers for diagnosis and disease evaluation of idiopathic pulmonary fibrosis," Chinese Medical Journal, vol. 136, no. 11, pp. 1278–1290, 2023.

[37] Y. Kunihiro, T. Matsumoto, T. Murakami, M. Shimokawa, H. Kamei, and N. Tanaka, "A quantitative analysis of long-term follow-up computed tomography of idiopathic pulmonary fibrosis: the correlation with the progression and prognosis," Acta Radiologica, p. 02841851231175252, 2023.

[38] J. Jacob, B. J. Bartholmai, S. Rajagopalan, M. Kokosi, A. Nair, and R. Karwoski, "Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures," European Respiratory Journal, vol. 49, no. 1, 2017.

[39] H. Nakagaw, Y. Nagatani, M. Takahashi, E. Ogawa, N. VanTho, and Y. Ryujin, "Quantitative CT analysis of honeycombing area in idiopathic pulmonary fibrosis: correlations with pulmonary function tests," European Journal of Radiology, vol. 85, no. 1, pp. 125–130, 2016.

[40] T. Handa, K. Tanizawa, T. Oguma, R. Uozumi, K. Watanabe, and N. Tanab, "Novel artificial intelligence-based technology for chest computed tomography analysis of idiopathic pulmonary fibrosis," Annals of the American Thoracic Society, vol. 19, no. 3, pp. 399–406, 2022.

[41] N. Su, F. Hou, W. Zheng, Z. Wu, and E. Linning, "Computed Tomography–Based Deep Learning Model for Assessing the Severity of Patients With Connective Tissue Disease–Associated Interstitial Lung Disease," Journal of computer assisted tomography, vol. 47, no. 5, pp. 738–745, 2023.

[42] W. Jianjian, G. Li, K. He, P. Li, L. Zhang, and R. Wang, "MCSC-UTNet: Honeycomb lung segmentation algorithm based on Separable Vision Transformer and context feature fusion," In Proceedings of the 2023 2nd Asia Conference on Algorithms, Computing and Machine Learning, 2023, pp. 488–494.

[43] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[44] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," IEEE transactions on medical imaging, vol. 39, no. 6, pp. 1856–1867, 2019.

[45] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, and Q. Tian , "Swin-unet: Unet-like pure transformer for medical image segmentation," in European conference on computer vision, Springer, 2022, pp. 205–218.

[46] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "Ds-transunet: Dual swin transformer u-net for medical image segmentation," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1–15, 2022.

[47] A Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, and B. Landman, “Unetr: Transformers for 3d medical image segmentation,” in: Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2022, pp. 574–584.

[48] G. Li, I. Yun, J. Kim, and J. Kim, “Dabnet: Depth-wise asymmetric bottleneck for real-time semantic segmentation,” arXiv preprint arXiv:1907.11357, 2019.

[49] T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, “Cgnet: A light-weight context guided network for semantic segmentation,” IEEE Transactions on Image Processing, vol. 30, pp. 1169–1179, 2020.

Efficient Deep Reinforcement Learning for Smart Buildings: Integrating Energy Storage Systems Through Advanced Energy Management Strategies

Artika Farhana[1], Nimmati Satheesh[2], Ramya M[3],
Janjhyam Venkata Naga Ramesh[4], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[5]
Dept. of Computer Science, Dair University College, Jazan University, Saudi Arabia[1]
Department of Computer Applications, PSNA College of Engineering and Technology, Dindigul[2]
Assistant Professor, Department of IT, Panimalar Engineering College, Chennai[3]
Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur Dist., Andhra Pradesh - 522302, India[4]
Faculty of Informatics and Computing, UniSZA University, Malaysia[5]

***Abstract*—This study presents a novel and workable approach to solving the critical issue of improving energy management in smart buildings. Using a large dataset from a seven-story office building in Bangkok, Thailand, our work introduces a novel approach that combines Deep Q-network (DQN) algorithms with energy storage models and cost optimization strategies. The suggested approach is intended to reduce operational expenses, improve the energy economic performance, and efficiently control peak demand. The energy storage model used in this research incorporates the use of the capabilities of advanced storage models in smart buildings, particularly lithium-ion batteries and supercapacitors. When the cost optimization approach is applied using linear programming, energy consumption costs are significantly reduced. Notably, our method outperforms current algorithms, specifically outperforming them, to show its effectiveness in smart building energy management by outperforming current algorithms, especially Genetic and Fuzzy Algorithms. In comparison to traditional methods, the DQN algorithm exhibits an impressive 8.6% reduction in Mean Square Error (MSE) and a 6.4% drop in Mean Absolute Error (MAE), making it a standout performer in the research through Python software. The results highlight the significance of optimizing DQN algorithm parameters for best outcomes, with a focus on adaptability to various properties of smart buildings. This investigation is novel because it integrates cost optimization, reinforcement learning, and energy storage. This results in a flexible and all-inclusive framework that can be used for effective and sustainable energy management in smart buildings.**

Keywords*—Deep q-network; cost optimization; smart building; energy management; peak demand

I. INTRODUCTION

Reinforcement learning integration with smart energy management is an attainable approach to improving energy system efficiency and optimizing consumption of energy. A subfield of machine learning called reinforcement learning (RL) trains agents to make decisions through interaction with their surroundings and feedback in the form of rewards or punishments. Applying reinforcement learning (RL) to smart energy management can help with complicated and dynamic decision-making problems [1]. RL algorithms perform well in settings where making decisions is dynamic and necessitates flexibility in response to shifting circumstances. With smart energy management, control techniques could be dynamically adjusted by RL for optimal energy utilization, taking into account variable elements such as weather patterns, user behaviors, and energy costs [2]. The algorithms have the potential to be utilized for precise load forecasting, energy demand pattern prediction, and energy-consuming device scheduling [3]. This lowers peak demand and maximizes the usage of energy resources. It may be used to enhance techniques for demand response. To take part in demand-side management programmes and gain incentives, agents can be trained to react to signals from utilities or energy suppliers and modify their patterns of energy usage [4]. The energy system can more easily incorporate renewable energy sources like wind and solar electricity thanks to these learning algorithms [5]. RL agents are able to optimize the use of renewable resources, balance supply and demand, and manage energy storage systems. It can maximize the control of distributed energy resources (DERs), such as solar panels, batteries, and electric cars, in microgrid settings [6]. Within the microgrid, agents can learn to balance the production and consumption of energy, increasing overall efficiency. RL is ideal for managing energy storage systems in an optimal way. Agents are able to acquire the best battery charging and discharging techniques, accounting for user preferences, grid circumstances, and power pricing [7]. Additionally, algorithms can assist energy management systems in adhering to legal and policy requirements [8]. Agents can be trained to make choices that adhere to regulatory requirements, energy efficiency standards, and other environmental considerations. Because these models can learn and adapt continuously, intelligent energy management systems can get better over time as they interact with their surroundings and gather input [9].

The implementation of machine learning techniques has led to significant improvements in energy management in smart buildings. Machine learning algorithms, especially those that employ supervised and unsupervised learning, provide insightful information and prediction powers that improve energy efficiency. To optimize lighting, other energy-consuming devices, HVAC systems, and occupancy trends, predictive models may be trained to examine historical data, weather patterns, and other pertinent variables [10]. Furthermore, anomaly detection algorithms can spot anomalous patterns in energy use, which allows for early intervention in the event of possible inefficiencies. Real-time adaptability is provided via reinforcement learning approaches, which dynamically modify control strategies in response to changing situations. The necessity for sizable labeled datasets, the interpretability of intricate ML models, and the possibility of biases in training data remain obstacles in spite of recent developments [11]. In order to protect against possible vulnerabilities, the deployment of ML models necessitates comprehensive consideration of cyber security measures [12]. Despite the encouraging outcomes of ML integration in smart buildings, these issues must be resolved to guarantee the stability and dependability of these systems in practical settings. Ongoing research and development initiatives are crucial to overcoming these obstacles and realizing the full potential of machine learning in smart building energy management as the field develops [13]. The capability of managing the intricate, nonlinear interactions present in energy systems is one of DRL's main advantages in smart buildings. Through constant environmental interaction and feedback in the form of incentives or penalties, DRL algorithms can acquire the best control rules for energy-intensive equipment such as lighting, HVAC systems, and other gadgets. Because of their adaptable nature, smart buildings can react quickly to changes in the weather, occupancy patterns, and energy prices, resulting in effective energy management. Deep learning integration makes it easier to uncover complex patterns from data, which leads to more precise forecasts and well-informed decision-making. Achieving a feasible deployment requires balancing the interpretability of the model with computing performance. The promise for significant gains in energy efficiency, cost savings, and sustainability continues to be a driving factor behind the development of intelligent building management systems as research into efficient DRL for smart buildings advances.

In the modern energy landscape, integrating energy storage technologies with sophisticated energy management methods is a critical first step towards improving efficiency, dependability, and sustainability. Because it may be used to store extra energy during times of surplus and release it during times of low generation or high demand, energy storage is essential for mitigating the intermittent nature of renewable energy sources. Advanced energy management strategies use complex algorithms, which frequently include machine learning and optimization approaches, to automatically regulate the cycles of energy storage systems' charging and discharging. These solutions optimize energy storage use by analyzing weather forecasts, historical data, and real-time demand patterns. This ensures that stored energy is strategically deployed to balance peak demand, minimize grid stress, and improve overall system resilience. Moreover, grid stability is enhanced and the integration of decentralized renewable energy resources is supported by the incorporation of energy storage into the energy management system. Despite these benefits, broad use will need to address issues including high upfront costs, the current generation of storage technologies' low energy density, and regulatory barriers. The secret to opening the door to a more robust and sustainable energy future lies in the smooth integration of energy storage systems through sophisticated energy management tactics, which will be made possible by ongoing technological and scientific developments. The following are the research study's main contributions,

- For the purpose of improving energy management in smart buildings, the study presents and uses Deep Reinforcement Learning algorithms. The study improves the system's capacity to make wise choices in a dynamic environment by utilizing DRL and taking into account variables like cost savings, peak demand reduction, and occupant comfort.

- The research provides a contribution by integrating a sophisticated energy storage model into the infrastructure of smart buildings. Modern technologies like lithium-ion batteries and supercapacitors are integrated into this concept and are arranged to store and harvest extra energy from renewable sources, improving sustainability and the economy.

- With the objective of minimizing overall energy consumption costs, the study presents a linear programming-based cost optimization model. This model offers a comprehensive approach to effective energy management by taking into account factors including electricity tariffs, operational costs, and potential fines for exceeding energy thresholds.

- A number of quantitative parameters are used in the study to evaluate the degree to which the suggested technique performs. These measures include the accuracy of the reinforcement learning model, cost savings, and percentage reductions in peak demand.

- The study emphasizes that the suggested tactics affect the lowering of peak demand. The results show that demand charges are significantly lower during peak hours, proving that the DRL strategy is effective in optimizing and modifying energy usage patterns.

- A comparative examination of the suggested DRL technique and a Genetic Algorithm (GA) approach is included in the paper. This comparison analysis shows that the DRL technique performs better in terms of peak load control and cost savings.

The paper's summary is given in Section I. Reviewing previous research, Section II highlights the gaps in the field's understanding of energy storage and management. The primary research question about the intricacies of smart building management is defined in Section III. Section IV presents the suggested technique. By comparing classifier performance, presenting empirical data in Section V, and examining conclusions and future research goals, Section VI demonstrates the importance of this research for smart building energy management.

II. Related Works

Elsisi et al. [14] presents a fresh and creative solution to the major problems associated with reducing energy usage and utilization in smart buildings, especially in the residential and commercial sectors. An impressive endeavor that places deep learning and the Internet of Things at the center of Industry 4.0 is the combination of these two technologies. A forward-thinking approach to effective energy management is demonstrated by the use of artificial intelligence tools in conjunction with the Internet of Things to share signals across machines and equipment. The paper's main contribution is the introduction of a people identification method based on deep learning that uses the YOLOv3 algorithm to maximize air conditioner performance and thereby cut energy usage. This method, which is based on precisely counting the people in a given space, allows for creative choices to be made for real-time air conditioner operational management in the context of smart buildings. The incorporation of the suggested system with an Internet of Things platform is also highlighted in the research. A dashboard receives internet-based updates on the number of people spotted and the condition of the air conditioners. This integration improves energy-related decision-making by offering insightful data on utilization trends and air conditioning usage. The simulation results that are reported in the research offer strong proof of the suggested approach's usefulness and effectiveness. The capacity of the deep learning-based identification algorithm to model extremely non-linear connections in data is demonstrated by its effective and accurate detection of the number of people in the designated region. The smooth broadcast of identification status on the dashboard of the IoT platform validates the system's usefulness. In conclusion, by utilizing deep learning and Internet of Things technology, the article significantly advances the subject of smart building power management. In addition to addressing the issues associated with energy consumption, the suggested method has potential uses in the remote control of a variety of controllable equipment. This study is positioned as a useful and innovative addition to the convergence of artificial intelligence, IoT, and energy conservation in smart buildings due to its integration of state-of-the-art technology and its encouraging simulation findings.

Shivam, Tzou, and Wu [15] presents a thorough machine learning-based multi-objective forecasting energy management plan for home grid-connected PV-battery hybrid systems. The hybrid approach under discussion combines an electric load in the form of a residential building, a bank of batteries for storing electricity, and a solar array. The suggested approach makes use of a three-tiered control framework: a dual forecasting framework based on residual causal dilation convolutional networks for generating electricity and electric load; a logical level for managing computational load and accuracy; and a multi-objective optimization for effective energy trade with the utility grid by means of battery charge scheduling. The prediction model exhibits precise one-step forward estimates for solar energy output and load, having been developed via a sliding window approach. The suggested energy management strategy is to minimize energy acquired through the utility grid, maximize the state of charge of the battery bank, and lower carbon dioxide emissions. Limits are placed on the highest amount of carbon dioxide that can be produced and the state of charge of battery banks. Using hourly power and load data, the approach is assessed under static as well as dynamic electricity pricing scenarios. The suggested dual prediction model has a high coefficient of predictability (93.08% for energy output and 97.25% for electrical load) according to simulation findings. The suggested prediction model shows substantial advances in accuracy when compared to naïve estimation, support vector machine, and artificial neural network (ANN) models. When combined with the sophisticated prediction model, the all-encompassing approach effectively controls more than half of the annual load demand, leading to notable decreases in carbon dioxide emissions and electricity costs when compared to residential structures with no hybrid energy systems or hybrid energy systems without an energy management plan. The research presents an organized and meticulous methodology, supported by comprehensive simulations, demonstrating its usefulness in incorporating machine learning into the predictive management of energy for home grid-connected PV-battery hybrid power systems.

Lan et al. [16] outlines a novel machine learning-based strategy for renewable microgrid energy management, with a special emphasis on a changeable structure made possible by remote tie and sectionalizing switches. The study notably uses sophisticated support vector machines (SVM) to model and estimate the hybrid electric vehicle (HEV) charging needs within the micro grid. To tackle the possible effects of HEV charge on the network, the study presents two discrete scenarios: intelligent and coordinated charging. A revolutionary modified optimization approach based on dragonflies addresses the complex nature of the issue formulation and provides a customized solution to the complicated problem. Also, a self-adaptive modification is suggested, which enables solutions to choose the modification strategy that best fits their particular situation. The effectiveness and suitability of the suggested strategy are shown by simulation findings on an IEEE microgrid evaluation system in both synchronized and autonomous charging scenarios. A high degree of precision is indicated by the mean absolute % inaccuracy of 0.978 for the anticipated total charge demand of HEVs. Moreover, the outcomes demonstrate a significant 2.5% decrease in the micro grid’s overall operating expenses when using the intelligent charging strategy in contrast to the coordinated method. The study advances the area by providing a thorough solution to the complex problems associated with controlling energy in renewable microgrids taking into account the needs of HEV charging. The relevance of the research is highlighted by the revolutionary reconfigurable structure, the personalized optimization strategy, and the use of modern machine learning techniques. The simulation findings validate the suggested approach's realistic deployment in renewable microgrid networks and offer compelling proof of its efficacy.

Syed et al. [17] focuses on the crucial component of dynamism estimating at the home level in smart constructions inside the larger framework of smart grid management of energy. Precise forecasts of energy usage in smart constructions are required for effective power generation and administration. The two primary phases of the suggested hybrid deep learning approach are model construction and data cleansing. Pre-processing techniques, such as adding lag values as extra features, are applied to raw data during the data-cleaning step. A hybrid deep learning architecture, comprising fully linked layers, unidirectional long-term short-term memory, and bidirectional LSTMs, is employed throughout the model-building stage. The objective of the model is to efficiently capture temporal relationships while maintaining high forecasting accuracy, low training time, and computing economy. The suggested model performs better than popular hybrid models like Convolutional Neural Networks, ConvLSTM, LSTM encoder-decoder frameworks, and stacking models, according to the evaluation of two benchmark energy consumption datasets. The suggested model achieves a mean percentage error in absolute terms of 2.00% for Case Study 1 and 3.71% for Case Study 2, indicating significant improvements. On the other hand, for the corresponding datasets, LSTM-based models produced greater MAPE readings of 7.80% and 5.099%. Furthermore, for the used energy consumption datasets, the suggested model shows promise in multi-step week-ahead everyday projections, exhibiting improvements in MAPE of 8.368% and 20.99% when compared to LSTM-based models. By presenting a unique hybrid deep learning model designed for household-level energy forecasts in Smart Buildings, the research makes a substantial contribution to the area. The thorough testing of the suggested method against well-known models and the documented increases in predicting accuracy highlight its possible applicability. This study offers useful insights for researchers and practitioners working in the fields of Smart Grid management of energy and Smart Buildings, especially about improving accuracy in forecasting household-level energy usage.

Han et al. [18] focuses on the potential of edge intelligence in the Internet of Things for green energy management, filling a major gap in the literature. The main goal is to provide a system based on deep learning for intelligent management of energy that can meet the needs of modern homes, businesses, and smart grids. The system attempts to forecast future short-term energy convention and enable effective statement among customers and energy providers. The paper's main contributions are the following: an innovative sequences learning-based energy forecasting mechanism, optimum normalization technique selection, and real-time energy management via devices at the edge interacting with a shared cloud-based data supervisory server. The lowest mistake rates and less temporal complexity are features of this forecasting system. According to the suggested architecture, edge devices connect in real time to a shared cloud server inside an IoT network, enabling efficient interactions across related smart grids and energy demand and response. To address the heterogeneous nature of electrical data, the study employs a number of preprocessing approaches. Next, an effective algorithm for decision-making is implemented for forecasting the immediate future on devices with limited resources. The efficiency of the suggested framework is demonstrated by extensive tests, which indicate a considerable decrease of 3.77 units for root MSE (RMSE) and 0.15 units for mean-square error (MSE) for commercial and residential datasets, respectively. The paper provides a significant addition to green energy conservation in the Internet of Things networks, especially when discussing edge intelligence. The suggested framework is a notable development in the field because of its useful applications in forecasting energy usage, improving communication, and lowering forecasting mistakes.

The reviewed studies collectively advance the field of energy management across diverse domains. According to one study, a novel solution to smart building energy management combines deep learning and IoT to maximize air conditioner efficiency by identifying individuals. A multi-objective forecasting strategy for residential grid-connected PV-battery hybrid systems is presented in another research. It uses a three-tiered control framework and achieves significant savings in power prices and carbon emissions. Within the field of renewable microgrid energy management, a study employing a modified optimization technique with support vector machines demonstrates efficacy in lowering total operating costs. Another study presents a hybrid deep learning approach with improved accuracy over conventional models, focusing on home-level energy forecasting in smart buildings. Finally, a study highlights the potential of edge intelligence in green energy management by putting forth a deep learning-based system that significantly lowers predicting errors and intelligently manages energy in Internet of Things networks. When taken as a whole, these research help to bridge the gap between cutting edge technologies such as IoT, machine learning, and deep learning by providing more effective and sustainable energy management techniques for a variety of applications.

III. Problem Statement

The effective management of energy in smart buildings is the issue that the literature addresses, with a special emphasis on the integration of energy storage systems using sophisticated energy management techniques. One of the current issues is making decisions effectively in a changing environment in order to maximize the utilization of energy, cut expenses, and improve overall operational efficiency. The research highlights the importance of using the Deep Q-Networks algorithm—a type of deep reinforcement learning (DRL)—as a solution to these problems. Emphasis is placed on the DQN algorithm's capacity to estimate and optimize the action-value function in a network of deep neural networks, demonstrating its efficacy in teaching appropriate energy management tactics. The suggested approach is to use DQN to help smart building managers make defensible choices about lighting, HVAC, and energy storage. Reducing peak demand, occupant discomfort, and costs significantly depends on the DQN algorithm's ability to manage intricate and dynamic interactions in the smart building environment. According to the literature, combining DQN with cost optimization methods and complex energy storage models results in a more effective and sophisticated method of managing energy in smart buildings [19].

IV. Proposed EMDQN Framework for Energy Management in Smart Building

The proposed energy management methodology for smart buildings follows a systematic sequence, beginning with the collection of a diverse dataset. Through Min-Max normalization, the dataset is pre-processed to ensure consistent scaling. The application of reinforcement learning evaluates the effects of centralized and decentralized model predictive controllers on peak demand and operational expenses. Simultaneously, an advanced energy storage model is introduced, utilizing lithium-ion batteries and supercapacitors to strategically capture and use excess energy. A cost optimization model, employing linear programming, aims to minimize overall energy consumption costs. The results and discussion section then analyzes the methodology's performance, focusing on metrics like peak demand reduction and cost-effectiveness. Deep Q-Networks (DQN) play a pivotal role in optimizing energy management, aligning decisions with objectives such as cost reduction, peak demand reduction, and occupant comfort. The entire process is illustrated through a flowchart, providing a comprehensive overview of the methodology's implementation from preprocessing to performance assessment. The entire methodology process is represented in Fig. 1.

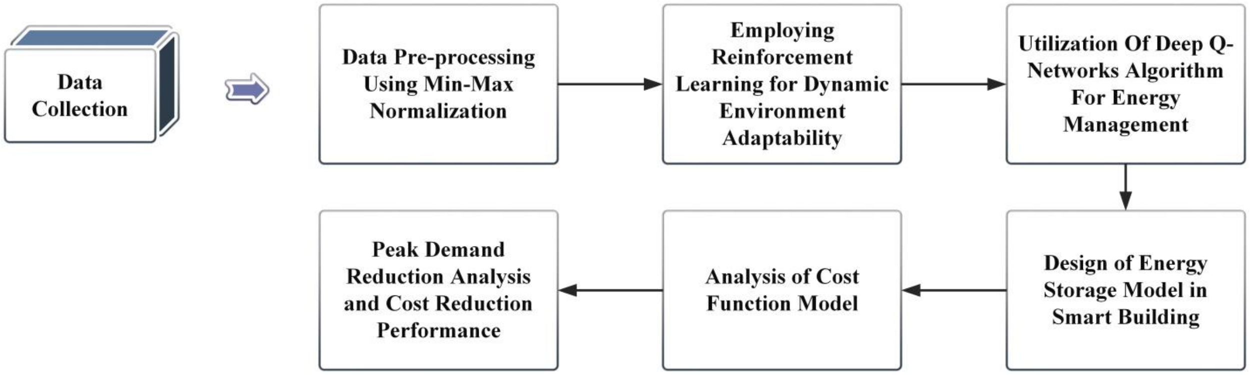

Data Collection
Data Pre-processing Using Min-Max Normalization
Employing Reinforcement Learning for Dynamic Environment Adaptability
Utilization Of Deep Q-Networks Algorithm For Energy Management
Design of Energy Storage Model in Smart Building
Analysis of Cost Function Model
Peak Demand Reduction Analysis and Cost Reduction Performance

Fig. 1. Proposed framework for energy management in smart building.

A. Data Collection

The valuable tool for study and advancement in the area of smart building energy management is the dataset gathered from Kaggle. This comprehensive dataset, which comes from a seven-story office building in Bangkok, Thailand, includes one-minute interval records of power usage and interior environmental measures. It covers the period from July 1, 2018, to December 31, 2019. The data on power use covers each of the building's 33 zones for plug loads, lights, and air conditioning systems. The collection is further enhanced by complementary interior environmental sensor data, which includes ambient light, relative humidity, and temperature readings for the same zones. The CU-BEMS dataset is distinct because it provides a thorough analysis of the electricity consumption at the building level, broken down by zone and floor, and it captures the functioning of important loads in commercial buildings. Numerous applications benefit greatly from such a dataset, such as multiple-level load forecasting, the development of indoor thermal models, the validation of building simulation models, the creation of demand response algorithms based on load types, and the application of reinforcement learning algorithms for multiple AC unit control. This dataset provides a solid foundation for optimizing energy consumption and storage techniques in smart buildings, which is in line with the goals of our planned study on smart buildings [20].

B. Data Pre-processing using Min-Max Normalization

The distributional properties of the original data are maintained by using Min-Max normalization. It provides a normalized representation that keeps the crucial data for training the model while scaling the values and preserving the connections and trends within the feature. All of the data points inside a feature are guaranteed to be scaled to a common range between 0 and 1 as a consequence of the normalization procedure. Outliers, which are data points significantly deviating from the majority, can distort the effectiveness of the subsequent analysis and modeling. Outliers can distort the uniform scaling intended by Min-Max normalization, leading to suboptimal model performance. By addressing outliers effectively during pre-processing, the dataset's integrity is preserved, ensuring that the subsequent energy management model is robust and reflective of the true underlying patterns in the smart building data. This uniform scaling is essential for preventing certain features from dominating the learning process due to their larger magnitudes is represented in Eq. (1).

$$X_i^m = \frac{X_i^m - X_{Min}^m}{X_{Max}^m - X_{Min}^n} \tag{1}$$

where, x_i^m is any value of a variable m; $X_{Max}^m$and X_{Min}^n are the maximum and the minimum values of that variable; $x_{i,scaled}^m$ is the value after scaling. By utilizing the Min-Max Scaler to normalize the input data, it is possible to prevent the issue where one characteristic overwhelms the others because of its larger range of values. If features are not normalized, the model could overweight the characteristics with higher values, thereby resulting in less than-ideal model performance. By scaling all characteristics to the same range, the Min-Max Scaler ensures that every feature has an equal impact on the model predictions.

C. Utilization of Reinforcement Learning in Dynamic Environment

The objective of Reinforcement Learning (RL), a machine learning technique, is to maximize a numerical reward while solving certain challenges in a predetermined environment. Numerous common and specialized engineering problems may be solved with this technology. Within the reinforcement learning paradigm, an agent engages in iterative interactions with its surroundings, picking up and applying certain behaviors based on the state of the environment. After that, the environment offers a reward in addition to its most recent condition, and so on, until the agent maximizes the total rewards obtained. The policy is often defined as the method by which an agent operates from a specific state. Finding the best course of action for the agent to maximize cumulative rewards in the given environment is the main objective. The reinforcement learning structure in Smart building is represented in Fig. 2.

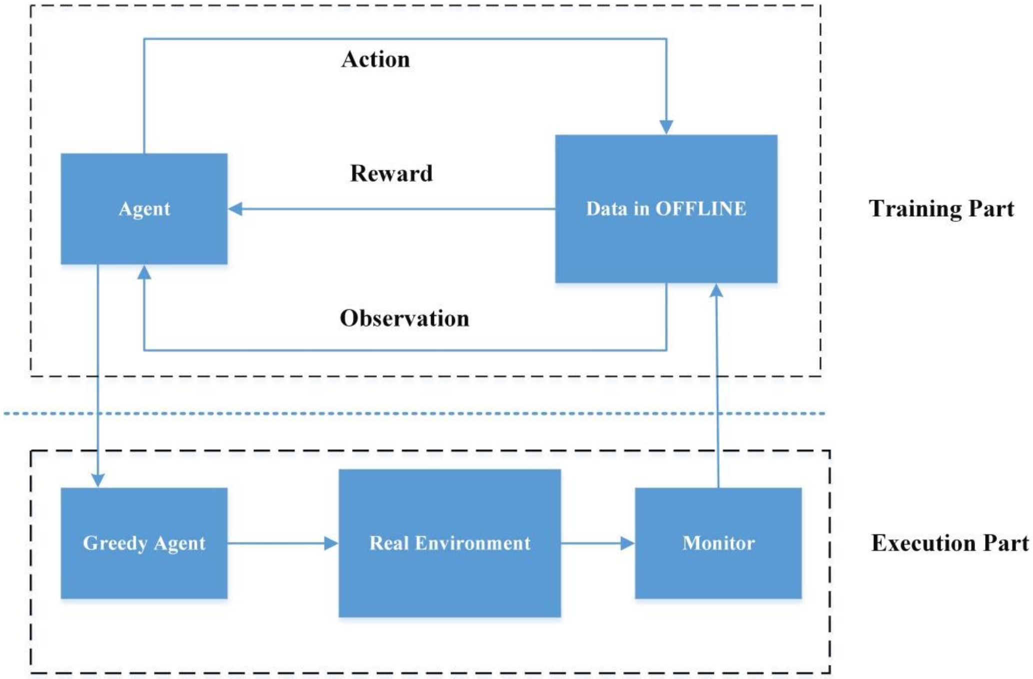


Fig. 2. Decision and control framework for reinforcement learning in smart energy management.

The environment in our research is assumed to be a Markov decision procedure, in which the agent's next state is determined only by its present state and the action it has selected, ignoring all other states and actions. The chosen value function for the inquiry is the Q-value, which is represented as$Q_p(a_x, b_x)$. The pairing of a state a_x and an action b_x at discrete time x is represented by this Q-value. The main goal of the agent is to maximize the Q-value at each time step. To find the best policy p in scenarios involving decision-making, Q-learning, a basic RL technique, is utilized. The Bellman equation is used in the Q-learning procedure to calculate and update the Q-value(a_x, b_x) is depicted in Eq. (2):

$$Q_p(a_x, b_x) = r(a_x, b_x) + \gamma max Q(a_x + 1, b_x + 1) \qquad (2)$$

In this case, the maximum discount future reward $\gamma max Q(a_x + 1, b_x + 1)$and the current reward $r(a_x, b_x)$add up to the ideal Q-value$Q_p(a_x, b_x)$. The relative relevance of present and future benefits is usually ascertained using a discounting factor $\gamma \in [0, 1]$. a smaller γ results in a more shortsighted agent that prioritizes immediate gains, whereas a bigger γ supports a more forward-looking strategy. To balance incentives for now and the future, the system operator can change the value of γ.

D. Deep Q-Networks (DQN) Algorithm for Energy Management Strategy

Deep Q-Networks is a reinforcement learning algorithm that combines deep learning with Q-learning to approximate and optimize the action-value function in a deep neural network. To determine the best practices for energy management in smart buildings, DQN is employed. In order to accomplish certain goals like cost reduction, peak demand reduction, and occupant comfort, it assists the system in making decisions about HVAC settings, lighting control, and energy storage measures. The Q-value represents the expected cumulative reward of taking a particular action in a given state is represented in Eq. (3)

$$Q_p(a_x, b_x) \leftarrow (1 - \alpha) \cdot Q_p(a_x, b_x) + \alpha \cdot (r + \gamma \cdot maxa' Q(a_x', b_x')) \qquad (3)$$

$Q_p(a_x, b_x)$ is the Q-value for state a_x and actionb_x. α is the learning rate, r is the immediate reward after taking action b_x in state a_x, γ is the discount factor, a_x' is the next state, b_x' is the next action. With parameters θ, a deep neural network approximates the Q-value. The Mean Squared Error between the goal Q-value and the predicted Q-value is the loss function used to train the network is presented in Eq. (4):

$$Loss(\theta) = E[(Q(a_x, b_x; \theta) - (r + \gamma \cdot maxa' Q(a_x', b_x'; \theta-)))^2] \qquad (4)$$

In this case, θ− stands for the parameters of a target network, which are updated with the online network's parameters, θ, on a regular basis. The input of deep neural network architecture is usually the state representation, and the architecture consists of fully linked layers. Each unit in the output layer represents the expected Q-value for a particular action, and there are as many units as there are potential actions. DQN frequently employs an epsilon-greedy strategy, in which the agent chooses an action (exploration) at probability ϵ and an action (exploitation) at probability 1−ϵ based on the maximum projected Q-value.

E. Energy Storage Model in Smart Building

For the purpose of maximizing cost-effectiveness, sustainability, and energy efficiency, an improved energy storage model is essential. In order to effectively store and manage energy and meet the changing demands of the building and its occupants, the model incorporates Super capacitors, lithium-ion batteries, when it comes to smart building energy management, this combination delivers clear benefits. Supercapacitors perform very well in cycles of fast charge and discharge, offering brief bursts of energy at times of peak demand and well balancing the intermittent nature of renewable energy sources. However, over longer periods of time, sustained and effective energy storage is guaranteed by lithium-ion batteries, which are renowned for their high energy density and dependability. Super capacitors, lithium-ion batteries, are the cutting-edge energy storage technologies are cleverly positioned to harvest extra energy produced from solar panels or other renewable energy sources during times of low demand. In order to ensure a steady and dependable power supply, this stored energy may subsequently be effectively used during periods of high demand or when renewable sources are insufficient. Predictive analytics and clever algorithms improve the system's responsiveness, allowing it to adjust to changing grid circumstances and energy needs. Smart building infrastructure is represented in Fig. 3.

The process for generating equations that represent the dynamics of energy storage, including the charging and discharging processes, is necessary to develop an energy storage model for a smart building. There is a single basic equation that determines the charge state (S_c) of the energy storage system over time can be expressed in Eq. (5):

$$S_c(x+1) = S_c\,(x) + c_x \frac{\gamma^{ch}\Delta\mathrm{x}}{Q_{storage}} p_{storage}^{charge}(x) + d_x \frac{\Delta\mathrm{x}}{\gamma^{disch} Q_{storage}} p_{storage}^{discharge}(x) \quad (5)$$

where, $p_{storage}^{charge}(x)$ and $p_{storage}^{discharge}(x)$ indicates the charging power and discharging power, $p_{storage}^{charge}(x) + d_x \geq 0$ and $p_{storage}^{discharge}(x) \leq 0$; γ^{ch} and γ^{disch}represents the discharging and charging effectiveness of the energy storage model ; The energy storage capacity and the time step duration are denoted by Δx. This formula is an example of a simplistic model; more intricate models may be developed by adding variables like round-trip efficiency, ageing effects, and temperature's effect on battery performance. These formulas enable the effective use of energy storage supplies while taking the dynamic nature of energy needs and external factors into account. They are crucial parts of a larger smart building energy management system.

F. Cost Optimization Model

The utilization of cost optimization models in smart building energy management often entails minimizing the total cost of energy consumption, taking into account variables such as power rates, operating expenses, and possible fines for over-energy thresholds. The following is an example of a popular formulation for this kind of model that uses linear programming is represented in Eq. (6):

$$C_x = \left(\frac{m_x - n_x}{2}|P_x| + \frac{m_x + n_x}{2}P_x\right),\ \forall\, x, \quad (6)$$

where, the energy purchasing and selling prices are represented by m_x and n_x. It is commonly recognised that the ESS's lifespan would be harmed by repeated charging or discharging. The Energy Storage System depreciation cost at time interval x is defined as follows to observe in Eq. (7),

$$C_x = \varphi(|c_x| + |d_x|) \quad (7)$$

where, c_x and d_x are charging and discharging power, and φ indicates the depreciation coefficient.

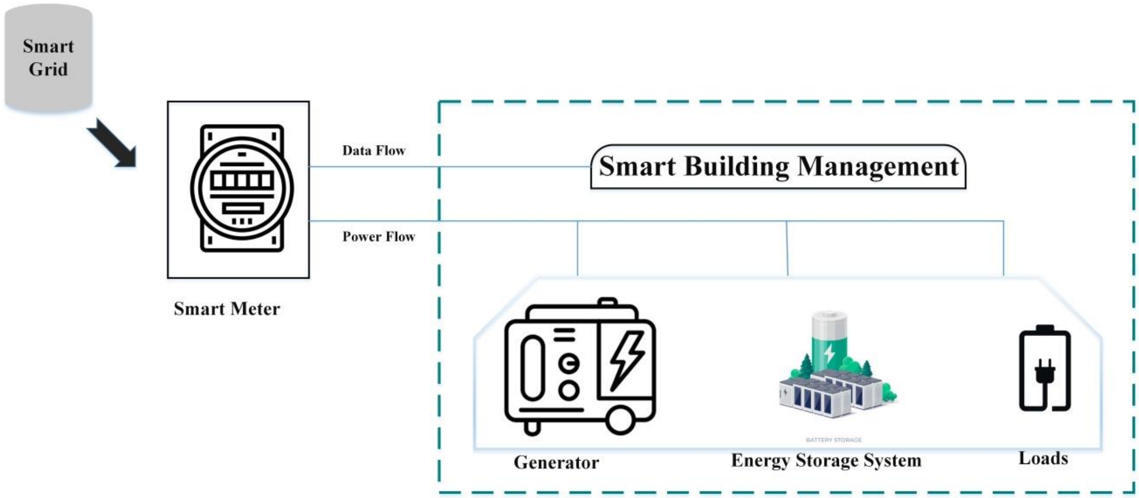


Fig. 3. Energy management system.

G. Reward

The reward function in a reinforcement learning setting typically represents the immediate benefit or cost associated with taking a particular action in a given state. In the context of energy management in smart buildings, the reward function can be designed to reflect the system's objectives. For instance, it might consider factors such as energy cost reduction, peak demand reduction, and occupant comfort. Here is a general form of the reward function, denoted as R in Eq. (8),

$$R(a_x, b_x) = r(a_x, b_x) + \gamma \cdot maxa'Q(a_x + 1, b_x + 1) \quad (8)$$

a_x and b_x represent the state and action at discrete time x. $r(a_x, b_x)$ is the immediate reward after taking action b_x in state ax. γ is the discount factor, determining the relative importance of present and future rewards. It is typically in the range [0, 1], where a smaller γ makes the agent more shortsighted, prioritizing immediate gains, while a larger γ supports a more forward-looking strategy.

Algorithm 1: EMDQN (Energy Management Deep-Q-Learning)

```
Input:
Raw dataset with power usage and environmental measures
Energy storage parameters: $\gamma^{ch}$, $\gamma^{disch}$, $Q_{storage}$, $\Delta x$, $\varphi$
Linear programming parameters: $m_x$, $n_x$
Output:
Trained DQN model
Optimized energy storage model
Cost-optimized smart building energy management system
def min max normalization(data):                    # Data Pre-processing:
  $X^m_{Min}$= min(data)
  $X^m_{Max}$ = max(data)
  Normalize data using (1)
  Return normalized data
    for t in range (max steps per episode):
      action = epsilon greedy(Q-network, state, $\epsilon$)
      Next state, reward = execute action(action)
      Store experience(state, action, reward, next state)
      Mini batch = sample mini batch()
      Update Q network(Q-network, target Q-network, mini batch)
      if t % $\tau$ == 0:
        Update target Q network(Q-network, target Q-network)
def DQN algorithm(Q-network, $\alpha$, $\gamma$):                  # Deep Q-Network (DQN) Algorithm
  for each episode:
    for each step:
      action = epsilon greedy(Q-network, state, $\epsilon$)
      execute action and observe reward and next state
      update Q-value using Eq. (3) and loss function Eq. (4)
def energy storage model($S_c$, $C_x$, $d_x$):               # Energy Storage Model
  Evaluate using (5)
def cost optimization model($m_x$, $n_x$, $C_x$, $d_x$, $\varphi$):        # Cost Optimization Model
   Evaluate using (6)
def reward function(r, $\gamma$, Q-value):                   # Reward Computation
  return r + $\gamma$ * max(Q-value)
Raw data = load dataset()                           # Main
Normalized data = min max normalization(raw data)
Q-network = initialize Q network()
Target Q-network = initialize Q network()
Reinforcement learning(Q-network, target Q-network)
Optimized energy storage model = energy storage model(parameters)
Optimized cost optimization model = cost optimization model(parameters)
```

V. Results and Discussion

The results section of the study focuses on evaluating the effectiveness and performance of the proposed methodology for efficient deep reinforcement learning in smart buildings, particularly concerning the integration of energy storage systems through advanced energy management strategies. Key metrics are employed to assess various aspects of the system's functionality, including cost optimization, peak demand reduction, occupant comfort, and overall energy efficiency. Quantitative measures, such as cost savings, peak demand reduction percentages, and the accuracy of the reinforcement learning model, will be reported. The results and discussion section receives inputs from the reinforcement learning model, energy storage model, and cost optimization model. The grid peak demand reduction performance and cost reduction performance analyses are part of the Results and Discussion section, providing insights into the effectiveness of the proposed methodology.

A. Cost Reduction Performance Analysis

Fig. 4 illustrates the breakdown of summertime electrical demand charges for the scenarios Without DQN, DQN-Temperature, DQN-Battery, and With DQN. It highlights interesting trends in peak demand management. August demand charges for the typical technique Without DQN are 1897 units; however, when the entire DQN strategy is integrated, these demand charges are significantly reduced to 1234 units, which is an outstanding decrease of almost 35.01%. Reductions are also achieved via the DQN-Battery and DQN-Temperature techniques, which have demand charges of 1698 and 1687 units, respectively. June and July reveal that DQN tactics work as well, with demand charges reduced by around 33.13% and 30.82%, respectively. On the other hand, September offers a special case with an unanticipated rise in demand fees for all DQN methods in contrast to the conventional model.

In Fig. 5, the DQN method produces the largest energy charge decrease in the winter, totaling 3897 units as opposed to 6785 units in the conventional Without DQN scenario, or a significant reduction of almost 42.61%. With a decrease in demand charges from 4587 units to 3298 units, or around 28.16% less, the reduction in demand charges is likewise noteworthy. The drop in cost is similarly notable in the summer. Compared to the conventional method, the With DQN technique helps to save around 16.99% in demand charges and 6.07% in energy charges.

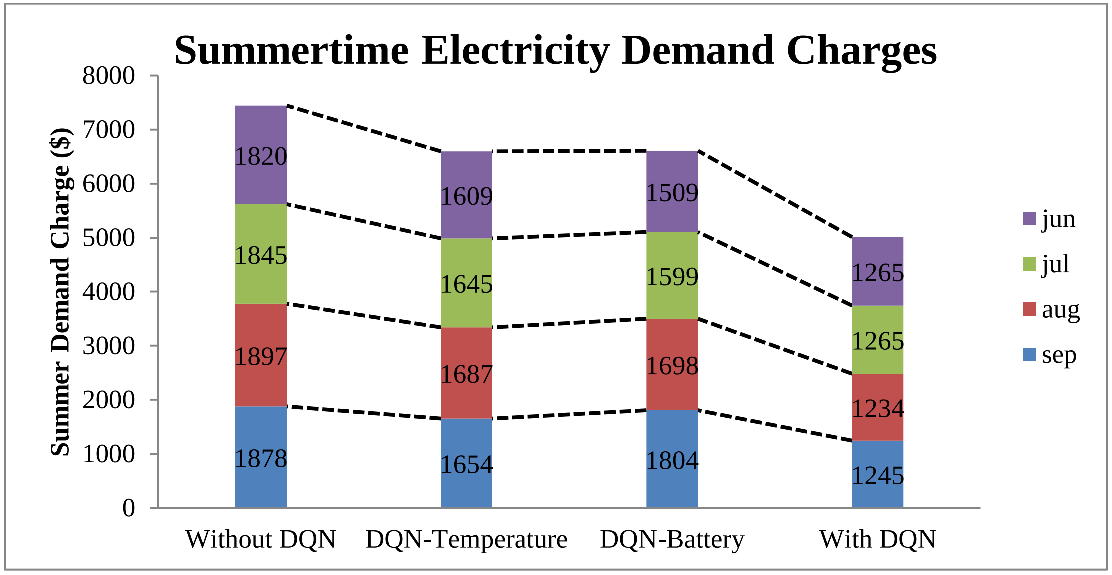


Fig. 4. Analysis of summertime electricity demand charges.

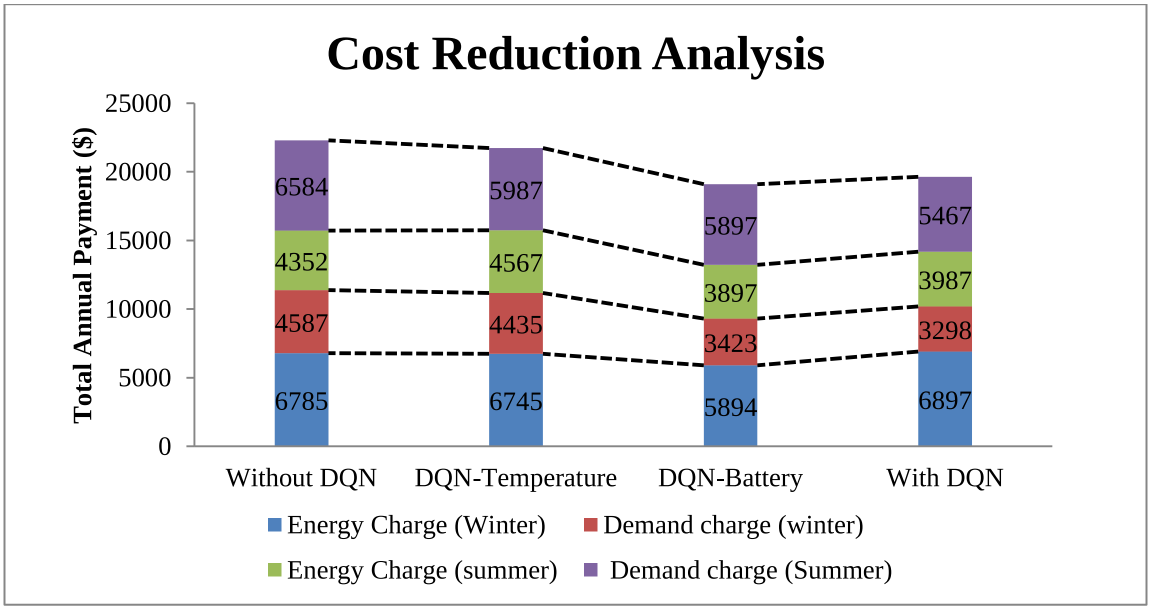


Fig. 5. Total annual cost reduction analysis.

The DQN techniques, especially the all-inclusive With DQN model, demonstrate how well they work to maximize energy use, tactically control peak demand, and eventually lower total operating expenses. This report highlights the potential financial benefits of using advanced energy management strategies and highlights DQN's contribution to the facility's large summer and winter cost savings.

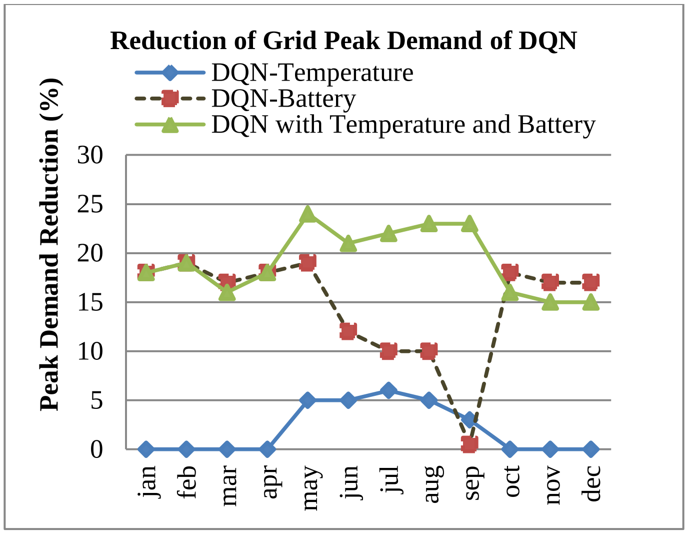


Fig. 6. Peak demand reduction evaluation.

Fig. 6 describes the observed peak demand reduction of 26.32% in May indicates a noteworthy beneficial impact of the adopted techniques, namely the DQN strategy, in the context of smart buildings. This significant decrease indicates that energy consumption habits were successfully adjusted and optimized during a month when demand is usually higher. The DQN strategy's ability to strategically manage and reallocate energy resources is demonstrated by its efficacy in May, which greatly enhances the building's total energy efficiency.

In Fig. 7, the rewards are received at various times when a reinforcement learning model is being trained or assessed. The rewards corresponding to each episode's progressive numbering show how well the model performed at each level. In this particular case, the incentives exhibit an increasing tendency as the number of episodes rises, going from an initial reward of -7.1 to -5 at episode 12,000 in this particular situation. The objective is to train the model to maximize its cumulative reward over episodes; the negative values indicate that the model is penalized for specific states or behaviors.

Table I compares the performance metrics of a proposed Deep Q-Network strategy with a Genetic Algorithm method for various situations of home energy management strategies. The peak load, power added to the grid, and power taken out of the grid are all represented by the Power figures.

The overall energy consumption and revenue/cost throughout time are tracked by cumulative energy and cumulative revenue/cost. The daily net cost is shown by the Net Cost. Comparing the Proposed DQN method to the GA approach, the percentage changes show that there is a reduction in peak load of 0.8918%, a fall in cumulative income of 3.4972%, a reduction in daily cost of 0.7647%, and an overall decrease in net cost of 6.7887%. These findings demonstrate the potential of the Proposed DQN method as an efficient home energy management strategy by indicating that it performs well in terms of peak load control and cost reduction.

B. Error Rate Evaluation

The mean square error (MSE) is a non-negative quantity, and its units are often the squared units of the original data. In regression analysis and other modeling tasks, it is commonly employed as an unbiased measure to assess how well a model with predictions or estimator is working. It is expressed in Eq. (9).

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(X_i - \widehat{X}_\iota\right)^2 \quad (9)$$

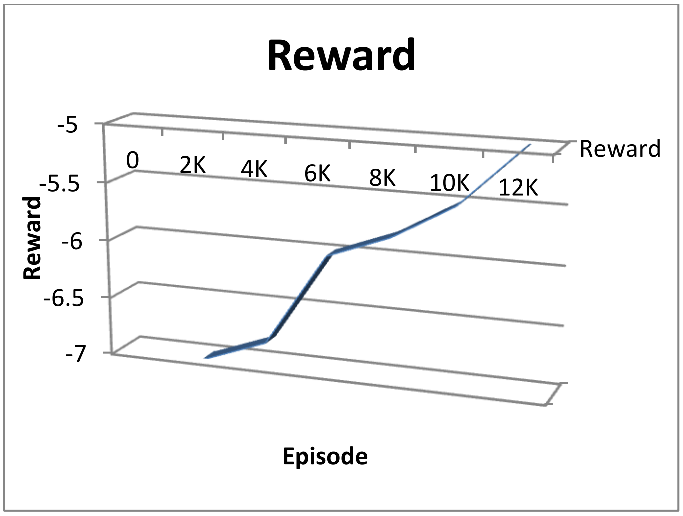


Fig. 7. Reward for proposed DQN algorithm.

TABLE I. COMPARISON OF THE RESULTS BETWEEN USING STRATEGIES GA AND PROPOSED DQN

Algorithm for Home Energy Management Strategy	Power (kW)	Cumulative Energy (kWh)		Cumulative Revenue/Cost (USD)		Net Cost (USD)
	Peak Load	Injected to Grid	Drawn from Grid	Revenue	Cost	Daily Cost
GA [21]	6.7894	53.0567	58.3765	2.7863	3.1243	0.3987
Proposed DQN	5.8976	53.0567	54.8793	2.7863	3.0596	0.3100
Increment (%)	-	-	-	-	-	-
Reduction (%)	0.8918	-	3.4972	-	0.7647	6.7887

where, m is the total amount of data points, X_i is the actual values and $\widehat{X_\iota}$ is the estimated values. MAE is used to indicate the quality of a prediction approach or estimator and is frequently given in the same units as the original data. It is expressed in Eq. (10).

$$MAE = \frac{1}{n}\sum_{i=1}^{m}\left|X_i - \widehat{X_\iota}\right| \tag{10}$$

where, m is the total amount of data points, X_i is the actual values and $\widehat{X_\iota}$ is the estimated values.

TABLE II. COMPARISON OF ERROR RATE

Energy Management Strategic Methods	Mean Square Error (%)	Mean Absolute Error (%)
Fuzzy Algorithm	14.5	15.87
Genetic Algorithm	12.6	14.43
Proposed DQN	8.6	6.4

The performance evaluation of three distinct energy management strategy methods—the Genetic Algorithm, the Proposed Deep Q-Network, and the Fuzzy Algorithm—is shown in Table II. Mean Absolute Error and Mean Square Error, both expressed in percentage terms, are the metrics used for evaluation. With a Mean Square Error of 14.5% and a Mean Absolute Error of 15.87%, the Fuzzy Algorithm exhibits a comparatively elevated degree of prediction mistakes. With a Mean Absolute Error of 14.43% and a Mean Square Error of 12.6%, the Genetic Algorithm performs more effectively. The proposed DQN performs much more effectively than both algorithms, with a Mean Square Error of 8.6% and a significantly lower Mean Absolute Error of 6.4%. Based on these findings, it appears that the Proposed DQN is a more promising method for optimizing energy systems than the Fuzzy and Genetic Algorithms in terms of prediction accuracy in energy management.

C. Discussion

The study's findings center on assessing the effectiveness and performance of the suggested approach for effective deep reinforcement learning in smart buildings, with a special emphasis on how improved energy management techniques integrate energy storage systems. Key performance indicators that are included in the evaluation include overall energy efficiency, peak demand reduction, cost optimization, and occupant comfort. Quantitative metrics are presented, including cost savings, percentages of peak demand decrease, and the accuracy of the reinforcement learning model. The results present the cost reduction study, which shows notable decreases in demand charges, energy charges, and overall operating expenditures in the summer and winter. The peak demand reduction evaluation shows a significant improvement in May due to the implemented approaches, with a decrease of 26.32%.

Furthermore, Methodology illustrates the incentives attained at various episodes throughout the model's training or assessment, exhibiting a progressive upward trend over time. The Proposed Deep Q-Network (DQN) method and the Genetic Algorithm [21] technique are thoroughly compared in Table I, which also highlights the benefits of the DQN strategy, which include a decrease in peak load, cumulative income, and total net cost. Taken as a whole, these measures highlight how well the suggested technique works to maximize energy consumption, cut expenses, and improve the overall performance of smart buildings. The study demonstrates the effectiveness of a systematic DRL framework for energy management in smart buildings. The proposed methodology integrates Min-Max normalization, reinforcement learning, and Deep Q-Networks to optimize decision-making processes. The integration of supercapacitors and lithium-ion batteries in the energy storage model is highlighted. This model aims to maximize cost-effectiveness, sustainability, and energy efficiency by strategically capturing and utilizing excess energy, addressing the intermittent nature of renewable energy sources.The study employs linear programming to develop a cost optimization model, aiming to minimize overall energy consumption costs. This model considers variables such as power rates, operating expenses, and potential fines for exceeding energy thresholds, providing a holistic approach to cost-effective energy management.

VI. CONCLUSION AND FUTURE WORKS

The energy management approach for smart buildings that is being presented demonstrates promise in terms of improving energy efficiency, cutting expenses, and efficiently controlling peak demand. It incorporates energy storage models, a Deep Q-network algorithm, and cost optimization techniques. Reinforcement learning is integrated into the system to enable dynamic adaptation to changing environmental circumstances and occupant demands. This optimizes decision-making processes related to energy storage measures, lighting management, and HVAC settings. When compared to conventional algorithms like fuzzy and genetic algorithms, the suggested DQN algorithm demonstrated notable improvements in cost reduction, peak demand management, and overall energy efficiency. The algorithm was developed on an extensive dataset from a seven-story office building in Bangkok, Thailand. Furthermore, the energy storage model improves the smart building's capacity for effective energy management and storage by cleverly using cutting-edge technology like lithium-ion batteries and supercapacitors. The linear programming-based cost optimization methodology helps to further reduce the overall cost of energy consumption. The effectiveness of the DRL technique may also be sensitive to the quality and representativeness of the dataset. Furthermore, the proposed energy storage model, while incorporating advanced technologies, simplifies the dynamics by using a basic equation, omitting factors such as round-trip efficiency and aging effects.

There are a number of directions to pursue in order to expand and enhance the suggested technique in future research. Initially, the flexibility and performance of the DQN algorithm might be improved by fine-tuning its hyperparameters and architecture to better fit the unique features of various smart buildings. Further optimizing the sustainability of the system might involve investigating the integration of renewable energy sources and implementing more sophisticated energy storage technology. Furthermore, given the dynamic nature of smart buildings, examining how online learning methods may be used to continually adjust to shifting circumstances and occupant behavior may help develop more adaptable and responsive energy management strategies. Moreover, broadening the scope of the cost optimization model to incorporate other variables like weather predictions, grid conditions, and dynamic pricing models may offer a more thorough method of reducing energy expenditures. All things considered, the approach that has been described provides a strong basis for smart building energy management. Research projects in the future can build on this foundation to tackle new problems and possibilities that arise in the quickly developing field of sustainable and energy-efficient building technology.

REFERENCES

[1] ‘A predictive and adaptive control strategy to optimize the management of integrated energy systems in buildings’, Energy Reports, vol. 8, pp. 1550–1567, Nov. 2022, doi: 10.1016/j.egyr.2021.12.058.

[2] W. J. Cole, J. D. Rhodes, W. Gorman, K. X. Perez, M. E. Webber, and T. F. Edgar, ‘Community-scale residential air conditioning control for effective grid management’, Applied Energy, vol. 130, pp. 428–436, Oct. 2014, doi: 10.1016/j.apenergy.2014.05.067.

[3] B. Celik, R. Roche, S. Suryanarayanan, D. Bouquain, and A. Miraoui, ‘Electric energy management in residential areas through coordination of multiple smart homes’, Renewable and Sustainable Energy Reviews, vol. 80, pp. 260–275, Dec. 2017, doi: 10.1016/j.rser.2017.05.118.

[4] S. Lee and D.-H. Choi, ‘Energy Management of Smart Home with Home Appliances, Energy Storage System and Electric Vehicle: A Hierarchical Deep Reinforcement Learning Approach’, Sensors, vol. 20, no. 7, Art. no. 7, Jan. 2020, doi: 10.3390/s20072157.

[5] J. Liu, X. Chen, H. Yang, and Y. Li, ‘Energy storage and management system design optimization for a photovoltaic integrated low-energy building’, Energy, vol. 190, p. 116424, Jan. 2020, doi: 10.1016/j.energy.2019.116424.

[6] ‘IEEE Xplore Full-Text PDF’: Accessed: Nov. 23, 2023. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9000577

[7] S. M. Hakimi and A. Hasankhani, ‘Intelligent energy management in off-grid smart buildings with energy interaction’, Journal of Cleaner Production, vol. 244, p. 118906, Jan. 2020, doi: 10.1016/j.jclepro.2019.118906.

[8] T. Gao and W. Lu, ‘Machine learning toward advanced energy storage devices and systems’, iScience, vol. 24, no. 1, p. 101936, Jan. 2021, doi: 10.1016/j.isci.2020.101936.

[9] W. J. Cole, J. D. Rhodes, W. Gorman, K. X. Perez, M. E. Webber, and T. F. Edgar, ‘Community-scale residential air conditioning control for effective grid management’, Applied Energy, vol. 130, pp. 428–436, Oct. 2014, doi: 10.1016/j.apenergy.2014.05.067.

[10] B. Lokeshgupta and S. Sivasubramani, ‘Multi-objective home energy management with battery energy storage systems’, Sustainable Cities and Society, vol. 47, p. 101458, May 2019, doi: 10.1016/j.scs.2019.101458.

[11] D. Coraci, S. Brandi, T. Hong, and A. Capozzoli, ‘Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings’, Applied Energy, vol. 333, p. 120598, Mar. 2023, doi: 10.1016/j.apenergy.2022.120598.

[12] Y. Ji, J. Wang, J. Xu, X. Fang, and H. Zhang, ‘Real-Time Energy Management of a Microgrid Using Deep Reinforcement Learning’, Energies, vol. 12, no. 12, Art. no. 12, Jan. 2019, doi: 10.3390/en12122291.

[13] A. T. Eseye and M. Lehtonen, ‘Short-term forecasting of heat demand of buildings for efficient and optimal energy management based on integrated machine learning models’, IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7743–7755, 2020.

[14] M. Elsisi, M.-Q. Tran, K. Mahmoud, M. Lehtonen, and M. M. Darwish, ‘Deep learning-based industry 4.0 and internet of things towards effective energy management for smart buildings’, Sensors, vol. 21, no. 4, p. 1038, 2021.

[15] K. Shivam, J.-C. Tzou, and S.-C. Wu, ‘A multi-objective predictive energy management strategy for residential grid-connected PV-battery hybrid systems based on machine learning technique’, Energy Conversion and Management, vol. 237, p. 114103, Jun. 2021, doi: 10.1016/j.enconman.2021.114103.

[16] T. Lan, K. Jermsittiparsert, S. T. Alrashood, M. Rezaei, L. Al-Ghussain, and M. A. Mohamed, ‘An Advanced Machine Learning Based Energy Management of Renewable Microgrids Considering Hybrid Electric Vehicles’ Charging Demand’, Energies, vol. 14, no. 3, Art. no. 3, Jan. 2021, doi: 10.3390/en14030569.

[17] D. Syed, H. Abu-Rub, A. Ghrayeb, and S. S. Refaat, ‘Household-level energy forecasting in smart buildings using a novel hybrid deep learning model’, IEEE Access, vol. 9, pp. 33498–33511, 2021.

[18] T. Han, K. Muhammad, T. Hussain, J. Lloret, and S. W. Baik, ‘An efficient deep learning framework for intelligent energy management in IoT networks’, IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3170–3179, 2020.

[19] X. Hou, J. Wang, T. Huang, T. Wang, and P. Wang, ‘Smart home energy management optimization method considering energy storage and electric vehicle’, IEEE access, vol. 7, pp. 144010–144020, 2019.

[20] ‘CU-BEMS, smart building energy and IAQ data’. Accessed: Nov. 23, 2023. [Online]. Available: https://www.kaggle.com/datasets/claytonmiller/cubems-smart-building-energy-and-iaq-data

[21] R. Liemthong, C. Srithapon, P. K. Ghosh, and R. Chatthaworn, ‘Home Energy Management Strategy-Based Meta-Heuristic Optimization for Electrical Energy Cost Minimization Considering TOU Tariffs’, Energies, vol. 15, no. 2, Art. no. 2, Jan. 2022, doi: 10.3390/en15020537.

Use of ANN, LSTM and CNN Classifiers for the New MSCC and BSCC Methods in the Detection of Parkinson's Disease by Voice Analysis

Miyara Mounia[1], Boualoulou Nouhaila[2], Nsiri Benayad[3], Belhoussine Drissi Taoufiq[4]

Computer Science and Systems Laboratory (LIS)-Faculty of Science Ain Chock, University Hassan II, Casablanca, Morocco[1]
Laboratory Electrical and Industrial Engineering-Information Processing-Informatics and Logistics (GEITIIL)-Faculty of Science Ain Chock, University Hassan II, Casablanca, Morocco[2, 4]
Research Center STIS, M2CS-National Higher School of Arts and Craft Rabat (ENSAM), Mohammed V University in Rabat, Rabat, Morocco[2, 3]

Abstract—**Parkinson's disease (PD) is a neurodegenerative condition that impacts a significant global population. The timely and precise identification of PD plays a pivotal role in facilitating early intervention and the efficient management of the condition. Recently, speech analysis has emerged as a promising non-invasive technique for the detection of PD due to its accessibility and ability to reveal subtle vocal biomarkers associated with the disease. This research introduces an innovative approach utilizing Short-Time Fourier Transform (STFT) to generate spectrograms, specifically Bark Spectrogram Cepstral Coefficients (BSCC) and Mel Spectrogram Cepstral Coefficients (MSCC). These coefficients are compared with traditional and well-known coefficients, namely Mel-Frequency Cepstral Coefficients (MFCC) and Bark Frequency Cepstral Coefficients (BFCC). To extract the most effective coefficients for Parkinson's disease detection, three robust classification techniques—Long Short-Term Memory neural networks (LSTM), Convolutional Neural Networks (CNN), and Artificial Neural Networks (ANN)—are employed. As a result, the BSCC and MSCC algorithms achieve a maximum accuracy rate of 90%, surpassing the accuracy of the traditional MFCC and BFCC coefficients. Therefore, these newly proposed coefficients prove to be more precise in diagnosing Parkinson's disease compared to the conventional MFCC and BFCC coefficients.**

Keywords—Parkinson's Disease (PD); Bark Spectrogram Cepstral Coefficients (BSCC); Mel Spectrogram Cepstral Coefficients (MSCC); Long-Term Memory Neural Networks (LSTM); Convolutional Neural Networks (CNN); Artificial Neural Networks (ANN)

I. INTRODUCTION

Parkinson's disease (PD) ranks among the most widespread and incapacitating neurodegenerative conditions, impacting millions of individuals across the globe. Named after the pioneering work of British physician James Parkinson, who first documented its clinical manifestations in 1817, Parkinson's disease has emerged as a profound challenge in modern medicine. Characterized by its progressive deterioration of motor function, PD also presents an intricate array of non-motor symptoms, encompassing sleep disturbances, cognitive impairments, emotional alterations, and autonomic dysfunctions.

This neurological condition is mainly characterized by the deterioration of dopaminergic neurons located in the substantia nigra part of the brain, resulting in a significant reduction in dopamine production. The repercussions of this disruption within the central nervous system manifest through a range of distinctive clinical symptoms, including muscle rigidity, resting tremors, bradykinesia (slowness of voluntary movements), and postural instability.

The timely and precise detection of PD is of paramount importance in providing timely and effective medical interventions. While current diagnostic methods primarily rely on clinical evaluation and specialized imaging techniques, an expanding body of research suggests that significant insights into the early detection of Parkinson's disease may be gleaned from analyzing the human voice. This hypothesis is rooted in the notion that subtle vocal changes, often imperceptible to the human ear, may serve as early indicators of the disease. Leveraging the advancements in machine learning and voice analysis technologies, researchers are increasingly exploring the potential of voice-based biomarkers for PD diagnosis.

In the academic literature, there is increasing attention to the application of speech-based techniques using both machine learning and deep learning for the detection of Parkinson's disease. Numerous research papers have investigated the application of machine learning methodologies, including SVM [1] – [4], KNN [5], [6] , DT [7] , and genetic algorithms [26] in this context. Simultaneously, the PD identification has been addressed through the employment of established convolutional neural network (CNN) architectures such as AlexNet, DenseNet, LSTM, SqueezeNet, VGG19, and others, as well as custom-designed CNN architectures developed by researchers for deep learning investigations [8], [9]. Notably, CNN architectures have demonstrated enhanced performance in the domain of feature extraction [8].

Moreover, CNN techniques have exhibited success in various research domains, encompassing tasks like ocean noise detection [10], COVID-19 detection via X-ray images [11]–[13], mammography image segmentation and classification [14], classification of environmental sounds [15], Alzheimer's disease detection [16], skin cancer diagnosis [17], identification of cartilage lesions [18], fatigue diagnosis based on heart sounds [19], diagnosis of joint disorders [20], premature retinopathy assessment [21], and even the diagnosis of idiopathic Parkinson's disease [22].

In this research, we explore the use of spectrogram-based techniques, specifically BSCC (Bark Spectral Cepstral Coefficients) and MSCC (Mel Spectral Cepstral Coefficients), to extract pertinent vocal features for disease classification. Our dataset, sourced from the SAKAR database, comprises 38 audio recordings, encompassing 18 patients diagnosed with PD and 20 healthy individuals. The stratification of the dataset allows for a comprehensive examination of the proposed methodologies in distinguishing between healthy and affected individuals.

This paper is structured as follows: Section I presents an extensive review of the background, contextualizing the significance of voice analysis in Parkinson's disease diagnosis. Section II delves into related work. Section III is about database. Section IV outlines the proposed methodology, elucidating the intricacies of the BSCC and MSCC techniques and their application in our study. Section V and Section VI encompasses the discussion of results, presenting findings, and insights derived from our experiments. Lastly, we summarize our main findings and their implications for further research and clinical applications in Section VII.

II. Related Work

The study of Mehmet Bilal et al. outlines a new approach for detecting PD based on voice signals using LSTM and pre-trained deep networks. The process involves four steps: noise reduction, mel-spectrogram extraction, deep feature extraction using pre-trained ResNet models, and classification with an LSTM model. Experiments conducted with the PC-GITA dataset, a widely used dataset, demonstrate superior classification performance compared to existing methods, emphasizing the importance of early diagnosis for speech-related Parkinson's symptoms, an accuracy of 98.61% was attained [23], for Gaffari selik et al. In their study, the research leverages advanced deep learning and machine learning techniques to diagnose PD using voice signal datasets from both PD patients and healthy individuals (PDO_Dataset and PD_Dataset). The study examines existing machine learning and CNN algorithms for PD diagnosis, conducting a comparative performance analysis. Furthermore, a novel approach named SkipConNet + RF, combining RF and CNN, is introduced for PD detection. SkipConNet extracts crucial features from voice signals and then employs the RF algorithm for estimation. This approach significantly enhances RF algorithm performance, achieving an improvement ranging from 3% to 17.19%. Remarkably, the SkipConNet + RF method achieves remarkable accuracy, with a 98.30% on the PDO_Dataset dataset and 99.11% success rate on the PD_Dataset dataset, showcasing its potential as a highly effective tool for PD diagnosis [24] .

The article of Quan et al. presents a novel deep learning approach for detecting PD through voice signals. The approach utilizes time-distributed 2D-CNNs to extract dynamic time series features and employs a 1D-CNN to capture dependencies between these features. The model's performance was evaluated on two databases. On Database-1, it outperformed traditional machine learning models, achieving accuracies of 81.6% for sustained vowel /a/ and 75.3% for a short Chinese sentence. On Database-2, the model attained up to 92% accuracy across various speech tasks, including reading sentences in Spanish. The model's learned time series features effectively captured variability and the reduced frequency range in PD sounds, crucial for diagnosis. Furthermore, the study highlights the significance of the low-frequency region in Mel-spectrograms for PD recognition from voice, surpassing the influence of the high-frequency region [25], For Karan et al. Their study addresses the use of speech as an early marker for PD detection, given its impact on several speech components. To overcome challenges related to non-stationarity and discontinuity in speech signals, the researchers introduce a novel feature called IMFCC based on empirical mode decomposition. The performance of these proposed IMFCC features is evaluated using two datasets, each comprising 25 Parkinson-affected individuals and 20 normal. The findings demonstrate that IMFCC features offer significantly improved classification accuracy in both datasets, with an impressive increase of 10–20% compared to MFCC features. This suggests the potential of IMFCC as a highly effective tool for PD identification through voice analysis [26]. Chen et al. employed an architecture that utilized the HHT and KNN algorithms. They extracted a total of 21 characteristics, consisting of 12 from each sound sample using the HHT algorithm and nine using the LPCC algorithm. Subsequently, these extracted characteristics were individually classified using KNN, DT and RF algorithms. The authors reported the best performance using the KNN, achieving an accuracy of 93.3% [27].

The study of Tasnas et al. Investigate the connection between speech dysfunction and PD, particularly focusing on novel dysphonia measures aimed at predicting PD symptom severity through speech signals. A sum of 132 dysphonia metrics was calculated based on sustained vowel sounds. These measures were then reduced to four parsimonious subsets using feature selection algorithms. These subsets were utilized for binary classification, employing both RF and SVM as statistical classifiers. The research leveraged a database with 263 samples from 43 subjects and demonstrated that these novel dysphonia measures can achieve remarkable results, with an overall classification accuracy of nearly 99% using only ten dysphonia features. The study highlights the complementarity of these newly proposed measures with existing algorithms, enhancing the classifiers' ability to distinguish PD subjects from healthy controls. These findings represent a significant advancement towards non-invasive diagnostic decision support for PD [28]. Yaman et al. utilized a freely accessible dataset from the UCI dataset platform, comprising 240 speech samples, with 40 from PD patients and 40 from healthy ones. Their approach involved augmenting the dataset's attributes through the application of a statistical pooling method. Subsequently, they derived weighted features employing the ReliefF technique. These weighted feature vectors were then subjected to classification using both SVM and KNN techniques, yielding a commendable 91.25% success rate with the SVM algorithm and 91.23% with the KNN method [29]. The paper of Kavita Bhatt et al. Highlights the significance of early PD detection in the elderly, emphasizing common symptoms such as dysarthria, tremors, cognitive changes, and muscle stiffness. The study proposes a Deep Neural Network algorithm using spectrograms generated by the Superlet Transform for PD detection from speech signals. The method achieved an impressive 96% on the ItalianPVS dataset and 92% overall accuracy on the PC-GITA dataset, outperforming state-of-the-art methods in PD detection from diverse speech data sources [30].

The research conducted by Mahboobeh and colleagues focuses on the demanding task of diagnosing PD earlier, particularly focusing on gender-specific differences in speech characteristics. A hybrid method is proposed, leveraging features' scores based on a two-dimensional projection. After removing gender-specific features, the approach employs Classification-Based Feature Score (CBFS) and Statistical-Based Feature Score (SBFS) to rank the remaining features. Resampling enhances feature selection stability. Various classifiers (KNN, NSVM, LSVM, RF, and NB) are applied, achieving 84% and 86% accuracy rates for women and men, respectively, with fewer selected features compared to prior studies. The results highlight feature commonalities despite gender differences and validate using an independent dataset for added robustness [31]. The proposed model comprises three key stages. Firstly, noise is eliminated from the signals using the DWT-EMD and EMD-DWT methods. Secondly, MFCC and GTCC features are extracted from the enhanced audio signals. The final step involves classification, where these features are input into CNN and LSTM models designed to capture sequential information from the extracted features. In the experimental phase, the study employs the PC-GITA and Sakar datasets, applying a ten-fold cross-validation method. Impressively, the highest classification accuracy for the PC-GITA dataset, the accuracy reaches 100% for EMD-DWT-GTCC-CNN and 96.55% for DWT-EMD-GTCC-CNN. For the Sakar dataset reaches 100% for both the EMD-DWT-GTCC-CNN and DWT-EMD-GTCC-CNN combinations. These findings underscore the effectiveness of GTCC features over MFCC in Parkinson's disease assessment. This research showcases a promising avenue for accurate and timely detection of PD through speech analysis, potentially improving the effectiveness of telemedicine-based diagnosis and monitoring systems [32].

III. Database

The study utilizes a dataset from a prior investigation, comprising 38 voice recordings. This dataset includes 20 individuals diagnosed with Parkinson's disease (PD) and 18 individuals classified as healthy controls. During these recording sessions, participants were specifically instructed to articulate the vowel 'a' using a standard microphone with a sampling frequency of 44,100 Hz, and the recordings were conducted on a desktop computer equipped with a 16-bit sound card. These voice recordings form the foundation of our analysis [33].

The proposed methodology involves the utilization of advanced feature extraction techniques as shown Fig. 1, including MFCC, MSCC, and BSCC, BFCC which will be detailed subsequently. These extracted acoustic features are then fed into powerful classification models, such as ANN, CNN and LSTM. The integration of these cutting-edge feature extraction methods and state-of-the-art classification algorithms forms the core of our approach, ultimately facilitating the accurate differentiation between individuals affected by PD and those in a healthy control group. The subsequent sections will provide a comprehensive explanation of each method and its role in our classification framework.

IV. Proposed method

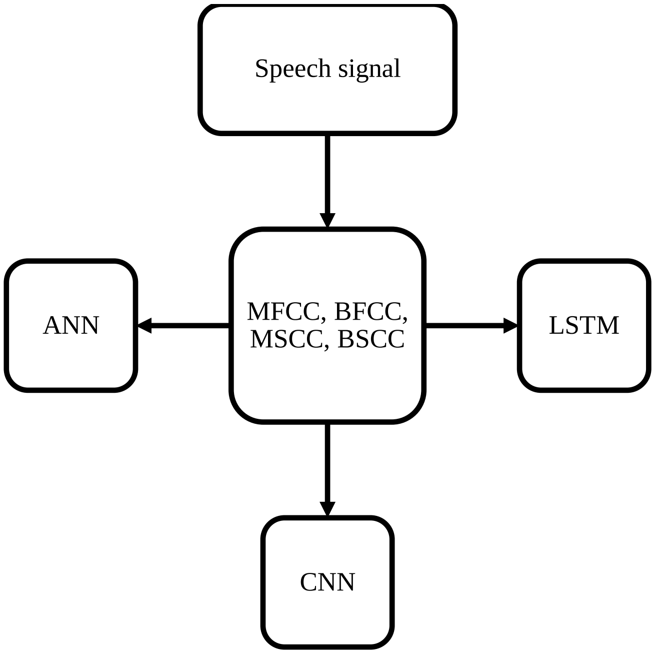


Fig. 1. The schematic for the proposed method.

A. MFCC

MFCC stands as a commonly employed method for extracting features in the realm of speech and audio signal processing, Fig. 2 depicts the different steps to follow in order to obtain the MFCC coefficients.

Pre-emphasis: The audio signal is pre-emphasized by applying a high-pass filter to amplify high-frequency components, as described this equation with k = 0.97.

$$\mathrm{H(z) = 1 - kz^{-1}} \tag{1}$$

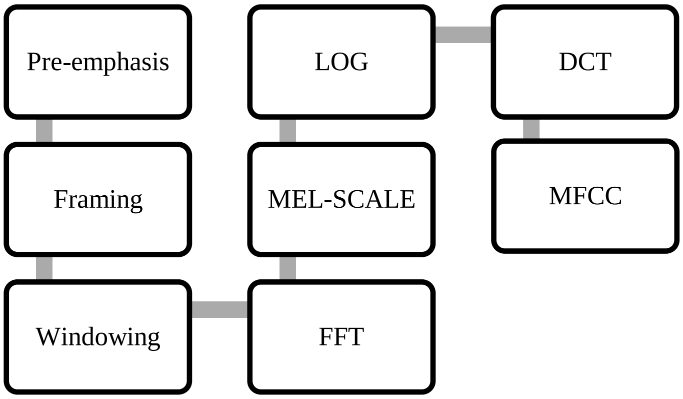


Fig. 2. The steps for calculating MFCC coefficients.

Framing: The signal with pre-emphasis undergoes segmentation into short, overlapping frames to capture temporal attributes.

Windowing: Each frame is windowed using the Hamming window function to prepare it for Fourier analysis.

$$w(n) = 0{,}54 - 0{,}46\cos\left(\frac{2\pi n}{N-1}\right) \qquad (2)$$

FFT (Fast Fourier Transform): The FFT is applied to each frame to convert it into the frequency domain.

$$S_n = \sum_{k=0}^{N-1} S_k e^{-j2\pi\frac{kn}{N}} \qquad (3)$$

Mel Scale: The resulting spectrum is transformed onto the Mel scale, which simulates human auditory perception.

$$\text{Mel} = 2595\log_{10}\left(1 + \frac{f}{700}\right) \qquad (4)$$

Discrete Cosine Transform (DCT): The DCT is used to decorrelate the Mel-scaled spectrum, reducing redundancy.

$$c_i = \sqrt{\frac{2}{N}}\sum_{j=1}^{M} m_j \cos\left(\frac{\pi i}{N}(j - 0{,}5)\right) \qquad (5)$$

MFCC Calculation: Finally, a subset of the DCT coefficients is selected as Mel-Frequency Cepstral Coefficients, which indicate the spectral properties of the audio signal.

B. BFCC

BFCC is similar to MFCC, but it uses the Bark scale instead of the Mel scale. In Fig. 3, the various stages required to acquire the BFCC coefficients are illustrated.

Pre-emphasis, Framing, Windowing, FFT: These initial steps are identical to those in MFCC.

Bark Scale: Instead of the Mel scale, the FFT results are transformed onto the Bark scale, which is another representation of auditory frequency perception.

$$\text{Bark}(f) = \frac{26{,}81f}{1960+f} + 0{,}53 \qquad (6)$$

DCT: The DCT is applied to the Bark-scaled spectrum to reduce redundancy and extract cepstral coefficients.

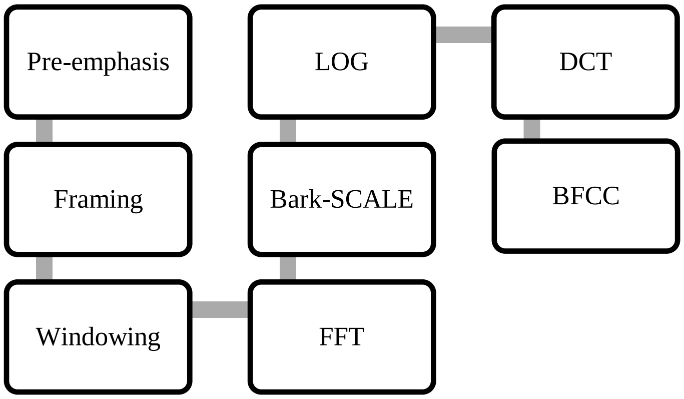


Fig. 3. The steps for calculating BFCC coefficients.

C. MSCC

MSCC is a variant of MFCC that directly uses the STFT instead of the FFT. The different procedures for obtaining the MSCC coefficients are presented in Fig. 4.

Pre-emphasis, Framing, Windowing: These steps remain the same as in MFCC.

STFT: Instead of the FFT, the STFT is used to attain a time-varying spectral representation for each frame.

Mel Scale and DCT: The Mel scale is implemented to the STFT results, followed by DCT, to calculate Mel Spectral Cepstral Coefficients.

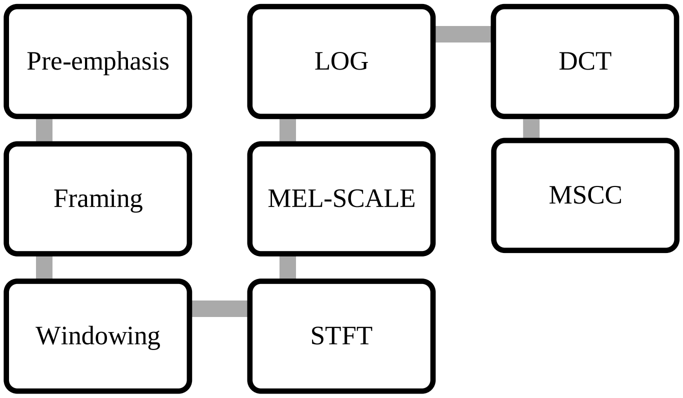


Fig. 4. The steps for calculating MSCC coefficients.

D. BSCC

BSCC is similar to MSCC, but it uses the Bark scale instead of the Mel scale in the final step. Fig. 5 outlines the sequential steps for deriving the BSCC coefficients.

Pre-emphasis, Framing, Windowing and STFT: These initial steps are identical to those in MSCC.

Bark Scale: Instead of the Mel scale, the STFT results are transformed onto the Bark scale, and then DCT is applied to extract Bark Spectral Cepstral Coefficients.

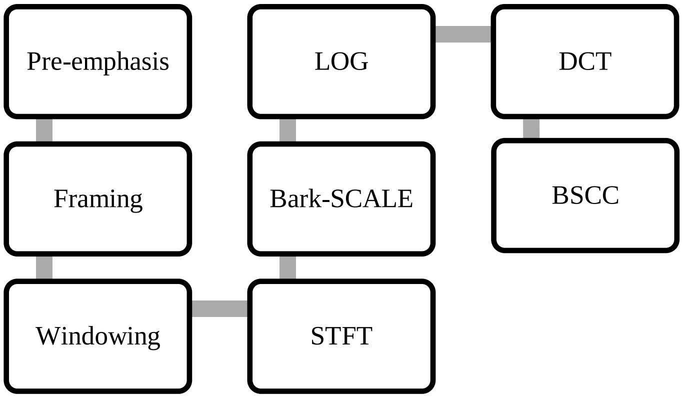


Fig. 5. The steps for calculating BSCC coefficients.

E. ANN

An Artificial Neural Network (ANN), modeled on the structure of biological neural networks, is a computational model composed of interconnected processing units, or neurons, organized into layers. These neurons process and transmit information through weighted connections. ANNs are designed for supervised learning and have the capacity to memorize complicated patterns and relationships within data. They include an input layer, one or more hidden layers, and an output layer. ANNs are widely employed for diversity of applications, including classification, regression, and function approximation.

F. *LSTM*

An LSTM is a form of RNN specifically created to tackle the challenge of the vanishing gradient issue encountered in conventional RNNs. LSTMs are notably effective for tasks that encompass sequential data. Time series data, or natural language processing. They incorporate specialized memory cells that can capture long-range dependencies in data, making them capable of retaining information over extended time intervals. LSTMs are crucial in applications such as speech recognition, and sentiment analysis, where understanding and modeling sequential patterns are essential.

G. *CNN*

A CNN is a deep learning structure mainly utilized for the analysis of images and spatial data. CNNs excel at automatically extracting hierarchical and spatial features from input data using convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply convolutional operations to scan and detect local patterns within the input, making CNNs highly effective in image classification, object detection, and image generation tasks. They have also found applications beyond image processing, including in natural language processing and reinforcement learning.

V. Result

In this section, the experimental outcomes are presented and the performance of the new approaches is assessed, MSCC and BSCC, alongside the BFCC and MFCC, in the context of PD detection through speech analysis. We also discuss the outcomes obtained using ANN, LSTM, and CNN for classification. Fig. 6 to Fig. 9 sequentially display the MFCC, MSCC, BFCC, and BSCC coefficients obtained for an individual diagnosed with Parkinson's disease.

The Fig. 6, 7, 8, and 9 displays the initial twelve coefficient values of MFCC, MSCC, BFCC, and BSCC, respectively. These coefficients encompass numerous frames that demand significant processing time for classification, hindering the accurate diagnostic decision-making process. To address this issue, the average value of these frames is computed to obtain the voiceprint and mitigate the processing burden.

The results of our experiments reveal varying levels of accuracy across different feature extraction methods and classification algorithms. For Mel Spectrogram-based Cepstral Coefficients (MSCC), we observed high accuracy rates, with ANN achieving 90% accuracy and CNN also reaching 90%, underscoring their effectiveness. LSTM, while slightly lower at 81%, still performed well with MSCC. Similarly, Bark Spectrogram-based Cepstral Coefficients (BSCC) exhibited strong accuracy, with ANN achieving 81% accuracy and LSTM and CNN both reaching 90%. In contrast, conventional BFCC and MFCC yielded comparatively lower accuracy rates, with ANN achieving 54% and 64%, respectively, and LSTM and CNN hovering around 80%. These findings suggest that MSCC and BSCC may offer superior feature extraction capabilities in the field of diagnosing PD, potentially revolutionizing the field with their higher accuracy rates when coupled with advanced classification techniques like CNN and LSTM.

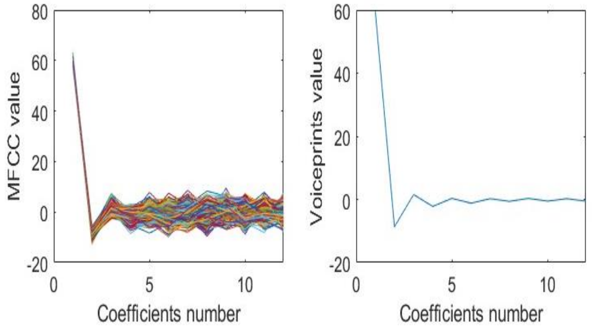


Fig. 6. The 12 MFCC coefficients for an individual with Parkinson's disease.

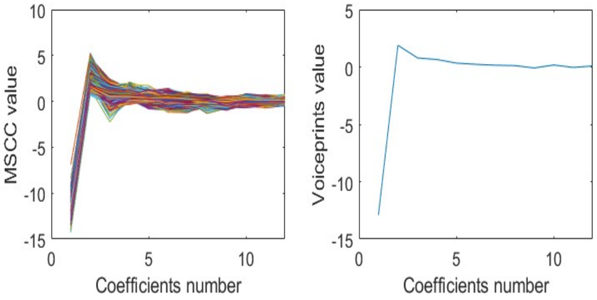


Fig. 7. The 12 MSCC coefficients for an individual with Parkinson's disease.

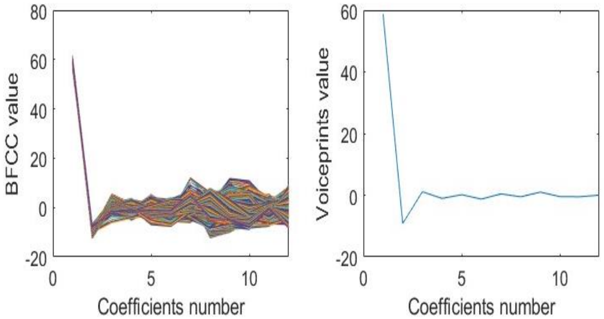


Fig. 8. The 12 BFCC coefficients for an individual with Parkinson's disease.

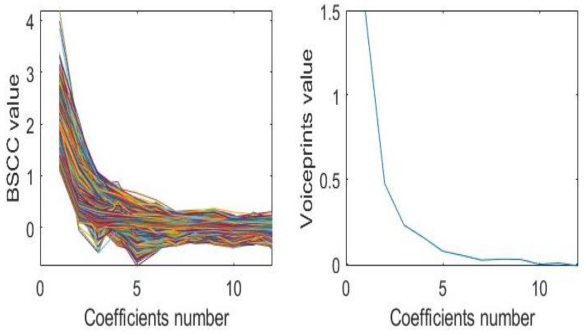


Fig. 9. The 12 BSCC coefficients for an individual with Parkinson's disease.

The achieved accuracy rates with MSCC and BSCC in combination with ANN, LSTM, and CNN are notable and indicate their promise as powerful features for PD detection. The consistent high accuracy rates of 90% with MSCC and BSCC when coupled with CNN highlight their efficacy in feature extraction. MSCC and BSCC capture spectral characteristics more effectively than traditional MFCC and BFCC, which may account for their superior performance. These findings suggest that MSCC and BSCC hold potential as valuable additions to the arsenal of voice-based PD detection methodologies.

In contrast, the results obtained using MFCC and BFCC, particularly with ANN, show comparatively lower accuracy rates of 54% and 64%, respectively. While these traditional coefficients have been widely used in voice analysis, our findings suggest that MSCC and BSCC surpass them in the context of PD detection. The improved accuracy achieved with MSCC and BSCC underscores their ability to capture nuanced vocal characteristics associated with PD more effectively.

TABLE I. RESULTS OBTAINED WITH THESE METHODS

	ANN			LSTM			CNN		
	Accuracy (%)	*Sensitivity (%)*	*Specificity (%)*	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
MSCC	90	60	87	81	50	50	90	50	100
BSCC	81	55	100	90	55	100	90	50	100
MFCC	54	100	100	81	66	100	81	71	100
BFCC	64	100	100	81	50	55	81	80	100

The accuracy rates obtained in this study are promising for the development of reliable voice-based PD detection systems. The utilization of MSCC and BSCC, combined with advanced deep learning techniques like CNN, offers a potential breakthrough in early PD diagnosis. These discoveries could hold noteworthy consequences for the development of non-invasive, cost-effective, and accessible PD screening tools, ultimately aiding in the timely intervention and management of this debilitating disease. Table I presents the results obtained for each of the MFCC, BFCC, MSCC, and BSCC coefficients using ANN, LSTM, and CNN.

In conclusion, our results suggest that MSCC and BSCC, in conjunction with CNN, represent a promising avenue for enhancing the accuracy of voice-based PD detection systems, potentially revolutionizing the way we diagnose and manage Parkinson's disease. Additional investigation and validation on larger datasets are warranted to confirm these findings and pave the way for practical clinical applications.

Three measures of performance were used in this study to evaluate the efficiency of classifiers on data sets: accuracy (see Eq. 7), sensitivity (see Eq. 8) and specificity (see Eq. 9). Accuracy is considered to be the percentage of precise results. Their definitions are as below:

$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TN+TP+FP+FN}} \quad (7)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad (8)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (9)$$

With:

Subjects without Parkinson's disease correctly categorized are True Negatives (TP).

Subjects with Parkinson's disease correctly categorized are True Positives (TN).

Subjects with Parkinson's disease incorrectly categorized are False Positives (FP).

Subjects without Parkinson's disease incorrectly categorized are False Negatives (FN).

VI. DISCUSSION

In the domain of diagnosing PD through vocal analysis, several methods have been proposed in the literature. This paper stands out, utilizing a combination of BFCC, MFCC, MSCC, and BSCC, combined with LSTM, ANN and CNN classifiers. Notably, the CNN-based method with MSCC ET BSCC achieved the highest accuracy of 90%, suggesting promising results for PD detection. Table II provides a comparison of this study with recently published articles.

TABLE II. COMPARISON WITH RECENT RESEARCH

Study	Dataset	Method	Accuracy
El-Hasnony et al.[34]	Sakar dataset	ANFIS + PSOGWO	87,5 %
Vasquez-Correa et al. [35]	Spanish datasets	CNN	89 %
Gunduz [36]	Sakar dataset	CNN	86,9 %
Sakar et al. [37]	Sakar dataset	SVM	86 %
Belhoussine et al. [38]	Sakar dataset	DWT-MFCC	86,84 %
Zayrit et al. [39]	Sakar dataset	SVM rbf SVM lin	81 % 79 %

In the domain of diagnosing PD through vocal analysis, several methods have been proposed in the literature. This paper stands out, utilizing a combination of BFCC, MFCC, MSCC, and BSCC, combined with LSTM, ANN and CNN classifiers. Notably, the CNN-based method with MSCC et BSCC achieved the highest accuracy of 90%, suggesting promising results for PD detection. Table II provides a comparison of this study with recently published articles.

Comparatively, previous research has explored various techniques for PD diagnosis through vocal analysis. El-Hasnony et al. [34] introduced a fog-based ANFIS+PSOGWO model, achieving an accuracy of 87.5% and outperforming other optimization methods Vasquez-Correa et al. [35] employed a CNN approach using STFT and continuous wavelet transform, achieving an accuracy of up to 89% in distinguishing PD patients from healthy speakers. Gunduz [36] proposed two CNN frameworks with deep feature extraction and achieved an accuracy of up to 86.90%. Sakar et al. [37] utilized tunable Q-factor wavelet transform for feature extraction and obtained an accuracy of up to 86%. Belhoussine et al. [38] focused on optimized wavelet selection in combination with MFCC features and SVM classification, achieving an accuracy of 86.84%. Finally, Zayrit et al. [39] used SVM classification with RBF kernel with Daubechies wavelet transform and MFCC features, obtaining an accuracy of 81%, but an accuracy of 79% by using SVM with linear Kernel.

Comparing these methods to the novel approach, which leverages MSCC, and BSCC with CNN classification to reach a 90% accuracy rate, it is evident that these methods outperform most of the previously mentioned techniques. This suggests that the combination of these advanced cepstral coefficients and CNN classification represent a significant advancement in the field of PD diagnosis through vocal analysis, potentially offering more accurate and reliable results. Further research and validation are necessary to confirm these findings and assess the practicality of implementing the proposed method in clinical settings.

VII. Conclusion

In summary, this document has delved into the exploration of Parkinson's disease through the lens of voice analysis, employing various feature extraction methods MFCC, MSCC, BFCC, and BSCC, coupled with a classification approach employing ANN, CNN, and LSTM. Notably, our findings consistently demonstrate that MFCC and BFCC methods consistently outperform others, achieving an impressive accuracy rate of 90%. These results underscore the potential of voice-based diagnostic tools in advancing our understanding and early identification of PD, highlighting the promising avenues for future research and clinical applications in this critical domaine. Subsequent investigations into the diagnosis of PD ailment aim to employ diverse neural network architectures and algorithms for feature selection on an expanded dataset.

References

[1] T. Zhang, Y. Zhang, H. Sun, and H. Shan, “Parkinson disease detection using energy direction features based on EMD from voice signal,” Biocybern Biomed Eng, vol. 41, no. 1, pp. 127–141, Jan. 2021, doi: 10.1016/j.bbe.2020.12.009.

[2] A. M. García et al., “Cognitive Determinants of Dysarthria in Parkinson’s Disease : An Automated Machine Learning Approach,” Movement Disorders, vol. 36, no. 12, pp. 2862–2873, Dec. 2021, doi: 10.1002/mds.28751.

[3] N. P. Narendra, B. Schuller, and P. Alku, “The Detection of Parkinson’s Disease From Speech Using Voice Source Information,” IEEE/ACM Trans Audio Speech Lang Process, vol. 29, pp. 1925–1936, 2021, doi: 10.1109/TASLP.2021.3078364.

[4] H. Gunduz, “An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson’s disease classification,” Biomed Signal Process Control, vol. 66, p. 102452, Apr. 2021, doi: 10.1016/j.bspc.2021.102452.

[5] Z. Soumaya, B. Taoufiq, N. Benayad, B. Achraf, and A. Ammoumou, “A hybrid method for the diagnosis and classifying parkinson’s patients based on time–frequency domain properties and K-nearest neighbor,” J Med Signals Sens, vol. 10, no. 1, p. 60, 2020, doi: 10.4103/jmss.JMSS_61_18.

[6] T. Tuncer, S. Dogan, and U. R. Acharya, “Automated detection of Parkinson’s disease using minimum average maximum tree and singular value decomposition method with vowels,” Biocybern Biomed Eng, vol. 40, no. 1, pp. 211–220, Jan. 2020, doi: 10.1016/j.bbe.2019.05.006.

[7] B. Nouhaila, B. D. Taoufiq, and N. Benayad, “An Intelligent Approach based on the Combination of the Discrete Wavelet Transform, Delta Delta MFCC for Parkinson’s Disease Diagnosis,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 4, pp. 562–571, 2022, doi: 10.14569/IJACSA.2022.0130466.

[8] M. B. Er, E. Isik, and I. Isik, “Parkinson’s detection based on combined CNN and LSTM using enhanced speech signals with Variational mode decomposition,” Biomed Signal Process Control, vol. 70, p. 103006, Sep. 2021, doi: 10.1016/j.bspc.2021.103006.

[9] L. Zahid et al., “A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson’s Disease,” IEEE Access, vol. 8, pp. 35482–35495, 2020, doi: 10.1109/ACCESS.2020.2974008.

[10] B. Mishachandar and S. Vairamuthu, “Diverse ocean noise classification using deep learning,” Applied Acoustics, vol. 181, p. 108141, Oct. 2021, doi: 10.1016/j.apacoust.2021.108141.

[11] S. Sheykhivand et al., “Developing an efficient deep neural network for automatic detection of COVID-19 using chest X-ray images,” Alexandria Engineering Journal, vol. 60, no. 3, pp. 2885–2903, Jun. 2021, doi: 10.1016/j.aej.2021.01.011.

[12] M. Aminu, N. A. Ahmad, and M. H. Mohd Noor, “Covid-19 detection via deep neural network and occlusion sensitivity maps,” Alexandria Engineering Journal, vol. 60, no. 5, pp. 4829–4855, Oct. 2021, doi: 10.1016/j.aej.2021.03.052.

[13] K. Shankar et al., “An optimal cascaded recurrent neural network for intelligent COVID-19 detection using Chest X-ray images,” Appl Soft Comput, vol. 113, p. 107878, Dec. 2021, doi: 10.1016/j.asoc.2021.107878.

[14] W. M. Salama and M. H. Aly, “Deep learning in mammography images segmentation and classification: Automated CNN approach,” Alexandria Engineering Journal, vol. 60, no. 5, pp. 4701–4709, Oct. 2021, doi: 10.1016/j.aej.2021.03.048.

[15] A. M. Tripathi and A. Mishra, “Self-supervised learning for Environmental Sound Classification,” Applied Acoustics, vol. 182, p. 108183, Nov. 2021, doi: 10.1016/j.apacoust.2021.108183.

[16] M. EL-Geneedy, H. E.-D. Moustafa, F. Khalifa, H. Khater, and E. AbdElhalim, “An MRI-based deep learning approach for accurate detection of Alzheimer’s disease,” Alexandria Engineering Journal, vol. 63, pp. 211–221, Jan. 2023, doi: 10.1016/j.aej.2022.07.062.

[17] A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[18] F. Liu et al., “Deep Learning Approach for Evaluating Knee MR Images: Achieving High Diagnostic Performance for Cartilage Lesion Detection,” Radiology, vol. 289, no. 1, pp. 160–169, Oct. 2018, doi: 10.1148/radiol.2018172986.

[19] C. Yin et al., "Diagnosis of exercise-induced cardiac fatigue based on deep learning and heart sounds," Applied Acoustics, vol. 197, p. 108900, Aug. 2022, doi: 10.1016/j.apacoust.2022.108900.

[20] U. Taşkıran and M. Çunkaş, "A deep learning based decision support system for diagnosis of Temporomandibular joint disorder," Applied Acoustics, vol. 182, p. 108292, Nov. 2021, doi: 10.1016/j.apacoust.2021.108292.

[21] J. M. Brown et al., "Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks," JAMA Ophthalmol, vol. 136, no. 7, p. 803, Jul. 2018, doi: 10.1001/jamaophthalmol.2018.1934.

[22] D. H. Shin et al., "Automated assessment of the substantia nigra on susceptibility map-weighted imaging using deep convolutional neural networks for diagnosis of Idiopathic Parkinson's disease," Parkinsonism Relat Disord, vol. 85, pp. 84–90, Apr. 2021, doi: 10.1016/j.parkreldis.2021.03.004.

[23] M. B. Er, E. Isik, and I. Isik, "Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with Variational mode decomposition," Biomed Signal Process Control, vol. 70, p. 103006, Sep. 2021, doi: 10.1016/j.bspc.2021.103006.

[24] G. Celik and E. Başaran, "Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of Parkinson's disease from speech signals," Applied Acoustics, vol. 211, p. 109476, Aug. 2023, doi: 10.1016/j.apacoust.2023.109476.

[25] C. Quan, K. Ren, Z. Luo, Z. Chen, and Y. Ling, "End-to-end deep learning approach for Parkinson's disease detection from speech signals," Biocybern Biomed Eng, vol. 42, no. 2, pp. 556–574, Apr. 2022, doi: 10.1016/j.bbe.2022.04.002.

[26] B. Karan, S. S. Sahu, and K. Mahto, "Parkinson disease prediction using intrinsic mode function based features from speech signal," Biocybern Biomed Eng, vol. 40, no. 1, pp. 249–264, Jan. 2020, doi: 10.1016/j.bbe.2019.05.005.

[27] L. Chen, C. Wang, J. Chen, Z. Xiang, and X. Hu, "Voice Disorder Identification by using Hilbert-Huang Transform (HHT) and K Nearest Neighbor (KNN)," Journal of Voice, vol. 35, no. 6, pp. 932.e1-932.e11, Nov. 2021, doi: 10.1016/j.jvoice.2020.03.009.

[28] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease," IEEE Trans Biomed Eng, vol. 59, no. 5, pp. 1264–1271, May 2012, doi: 10.1109/TBME.2012.2183367.

[29] O. Yaman, F. Ertam, and T. Tuncer, "Automated Parkinson's disease recognition based on statistical pooling method using acoustic features," Med Hypotheses, vol. 135, p. 109483, Feb. 2020, doi: 10.1016/j.mehy.2019.109483.

[30] K. Bhatt, N. Jayanthi, and M. Kumar, "High-resolution superlet transform based techniques for Parkinson's disease detection using speech signal," Applied Acoustics, vol. 214, p. 109657, Nov. 2023, doi: 10.1016/j.apacoust.2023.109657.

[31] M. Hasanzadeh and H. Mahmoodian, "A novel hybrid method for feature selection based on gender analysis for early Parkinson's disease diagnosis using speech analysis," Applied Acoustics, vol. 211, p. 109561, Aug. 2023, doi: 10.1016/j.apacoust.2023.109561.

[32] N. Boualoulou, T. Belhoussine Drissi, and B. Nsiri, "Cnn and Lstm for the Classification of Parkinson's Disease Based on the GTCC and MFCC," Applied Computer Science, vol. 19, no. 2, pp. 1–24, Jun. 2023, doi: 10.35784/acs-2023-11.

[33] B. E. Sakar et al., "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings," IEEE J Biomed Health Inform, vol. 17, no. 4, pp. 828–834, Jul. 2013, doi: 10.1109/JBHI.2013.2245674.

[34] I. M. El-Hasnony, S. I. Barakat, and R. R. Mostafa, "Optimized ANFIS Model Using Hybrid Metaheuristic Algorithms for Parkinson's Disease Prediction in IoT Environment," IEEE Access, vol. 8, pp. 119252–119270, 2020, doi: 10.1109/ACCESS.2020.3005614.

[35] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Convolutional neural network to model articulation impairments in patients with Parkinson's disease," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, International Speech Communication Association, 2017, pp. 314–318. doi: 10.21437/Interspeech.2017-1078.

[36] H. Gunduz, "Deep Learning-Based Parkinson's Disease Classification Using Vocal Feature Sets," IEEE Access, vol. 7, pp. 115540–115551, 2019, doi: 10.1109/ACCESS.2019.2936564.

[37] C. O. Sakar et al., "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," Appl Soft Comput, vol. 74, pp. 255–263, Jan. 2019, doi: 10.1016/j.asoc.2018.10.022.

[38] T. B. DRISSI, S. ZAYRIT, B. NSIRI, and A. AMMOUMMOU, "Diagnosis of Parkinson's Disease based on Wavelet Transform and Mel Frequency Cepstral Coefficients," International Journal of Advanced Computer Science and Applications, vol. 10, no. 3, 2019, doi: 10.14569/IJACSA.2019.0100315.

[39] Z. Soumaya, B. D. Taoufiq, N. Benayad, B Achraf, & A. Ammoumou, (2020). A hybrid method for the diagnosis and classifying parkinson's patients based on time–frequency domain properties and K-nearest neighbor. Journal of medical signals and sensors, 10(1), 60.

Artificial Intelligence for Confidential Information Sharing Based on Knowledge-Based System

Bouchra Boulahiat, Salima Trichni, Mohammed Bougrine, Fouzia Omary
Faculty of Sciences of Rabat-Computer Science Department, Mohammed V University in Rabat, Rabat – Morocco

***Abstract*—Ensuring the security of sensitive data and protecting user privacy remains one of the most significant challenges in our contemporary landscape. Organizations Companies cannot adopt a new technology without reassurance regarding data confidentiality. To address these challenges, we present an innovative system that draws upon extensive knowledge and expertise in the field of cryptography, especially in encryption methods. This system tailors its strategies to align with specific scenarios, prioritizing data confidentiality. Our solution is based on one of the Artificial Intelligence techniques, which is Knowledge-Based Systems (KBS) and extends the intelligent encryption methods from our previous research. However, this new system has taken a novel approach by reconfiguring this within KBS architecture. We have introduced additional technical components, including knowledge bases, an inference engine, and the Nearest Neighbor (NN) search algorithm. As a result, this revised architecture not only enhances security and system performance but also showcases improved maintainability and scalability.**

Keywords—IT security; cryptography; confidentiality; Knowledge-Based system; artificial intelligence

I. INTRODUCTION

IT security is an immensely influential domain that plays a pivotal role in shaping the trajectory of the broader IT landscape. The digital realm evolves exponentially, consistently introducing innovative concepts to simplify our lives and enhance our daily experiences. However, this technological evolution, when it intersects with fundamental human values and the need to uphold privacy, is stripped of its scientific significance and weight.

Security considerations have remained a persistent challenge from the inception of computer systems to the present day. The paramount question that incessantly prevails is that of data security. Initially, this question was primarily concerned with data confidentiality. Nevertheless, the emergence of the internet ushered in a multitude of additional requirements, including the assurance of data integrity, authenticity, availability, and non-repudiation. This evolution gave rise to various cryptographic primitives, encompassing both symmetric and asymmetric encryption algorithms, hash functions, digital signatures, and digital certificates.

However, the pace of creating new cryptographic tools tailored to evolving technological needs has significantly slowed down. Contemporary enterprises tend to gravitate towards well-established algorithms celebrated for their enduring performance and resilience over the years, as opposed to seeking novel cryptographic systems. Consequently, to adopt a new technology, these companies must establish a robust security policy to accommodate the requisites of the technology. This policy relies on a set of cryptographic primitives, along with additional methods for managing access and infrastructure.

Nonetheless, modern IT systems are increasingly characterized by their diversity, interactivity, and the need to interact with a continually expanding network of third parties. These third parties may not necessarily adhere to the same security standards as other established partners. This presents a dilemma of rigidity in security protocols: either rigidly enforces identical security policies across all interactions, a measure that may limit user participation, or continuously adapt the system to accommodate new customer requirements, an approach that can prove costly and impact existing customers. Existing methods might be too inflexible in their security protocols, making it challenging to adapt or align with varying security standards of third-party networks. This inflexibility could hinder effective collaboration with diverse partners.

To address this predicament, we propose a novel encryption model designed to flexibly align with the specific demands of each situation. By implementing this mediation, we can make informed decisions regarding the most suitable encryption algorithm.

The reasons that make the proposed encryption model suitable for addressing the challenges mentioned in the paragraph:

1) Flexibility and adaptability: The encryption model is designed to be flexible, allowing it to adapt to varying security standards and requirements specific to each situation.

2) Tailored security measures: It can accommodate diverse IT systems, interact with a broad network of third parties, and align security measures accordingly without compromising overall protection.

3) Customization for Different Situations: It enables customization based on individual demands, ensuring that security protocols can be adjusted as needed, even when dealing with partners who might not adhere to standard security practices.

4) Cost-effective solution: The model aims to balance the need for security enhancements with cost considerations, ensuring that implementing and adapting security measures remains economically viable.

5) Minimized Impact on Existing Customers: By offering flexibility, it minimizes the impact on existing customers while accommodating new requirements, avoiding disruptions in service or user experience.

Overall, the proposed encryption model aims to provide a versatile and adaptive solution that addresses security challenges in modern IT systems without excessively limiting user engagement or imposing exorbitant costs.

In the following sections, we will delve into the intricacies of this approach. We will commence by outlining our motivation and the prior research undertaken in this domain. Subsequently, we will expound upon the proposed solution, detailing its underlying principles and the various modules that constitute it, starting from data classification and extending to its inference engine. Lastly, we will elucidate the application and experimental aspects of this approach, which have been tested against a diverse array of the most renowned encryption algorithms.

A. Background

Reviewing the literature alongside various research endeavors concerning established encryption algorithms [1] [2] [3], and drawing from our own hands-on experience with these algorithms [4] [5] [6] [11], it becomes apparent that the level of security within each encryption system remains far from constant. This security profile tends to fluctuate from one study to another. What one might deem the ideal algorithm, such as AES, in one scenario, may not hold the same status in another [7]. In essence, the performance profile of each algorithm undergoes variations contingent upon the specific context and the environment in which it operates.

To illustrate this, let us consider a comparative study mentioned in [8]. This study assessed the performance of symmetrical encryption algorithms, namely DES, AES, and Blowfish, in the context of processing images. The study involved several images of varying sizes, each accompanied by their respective histograms. By comparing the encryption and decryption times of these algorithms, the research presented a comparative diagram in “Fig. 1”, as demonstrated below:

Examining “Fig. 1”, the following conclusions can be drawn:

- DES excels when dealing with small images.
- Blowfish stands out for its efficiency in processing large images in terms of cipher time.
- Blowfish and AES exhibit nearly identical performance levels during decryption.

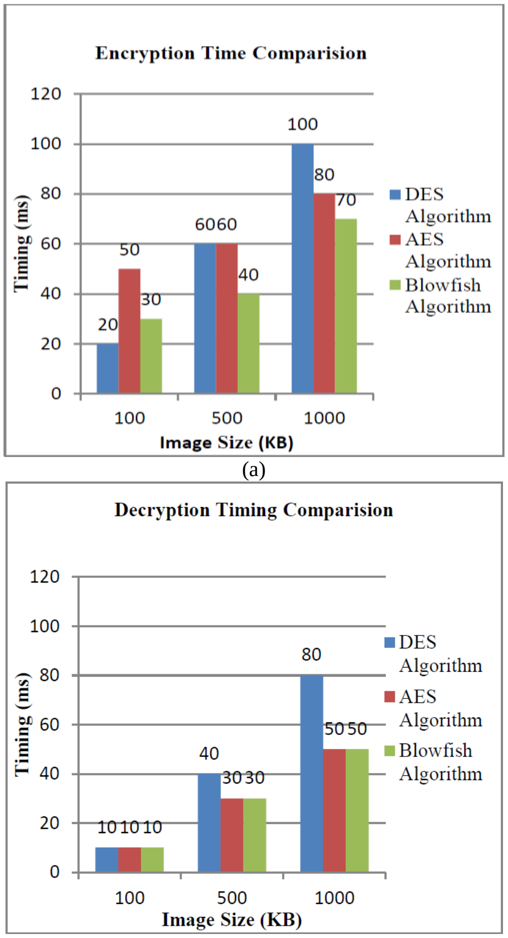
In a different study, which has also contributed to inspiring our approach and addresses various constraints, data types were considered [9]. This research explored the performance of encryption algorithms on mobile platforms, focusing on Triple DES, AES, and other methods based on elliptic curve arithmetic (ECC). The study sought to evaluate their performance across different Android platforms, including:

- Acer Iconia Tab A511
- Samsung Galaxy S4 i9505
- LG P500 Optimus One

Within each environment, numerous tests were conducted to assess algorithm performance based on the storage type, whether on an SD card or internal storage. The results from this study diverge from the previous one, highlighting the following insights:

- On the Samsung Galaxy S4 i9505, AES and DESede performed similarly, while ECC is not recommended.
- On the LG P500, AES remained the superior choice.
- On the Acer Iconia Tab A511, ECC and AES demonstrated comparable performance for smaller input file sizes, although ECC is not recommended for handling larger text files.

Hence, relying solely on a single encryption system for all communications could potentially undermine both security and system performance.



(a)

(b)

Fig. 1. Comparison between the timing of image encryption algorithms (a) and image decryption algorithms (b) using different image sizes.

In our prior works [5] [10], we devised a system that enables the categorization of various encryption scenarios within a decision-making database structured using a star modeling, comprising a central fact table and multiple dimensions. This system involves extracting the characteristics of each communication and employs data warehousing techniques to determine the most secure encryption algorithm for a case similar to or closely aligned with our own.

Building upon this foundational principle, this work introduces a substantial overhaul of the system, adopting a novel architecture based on a Knowledge-Based System (KBS). This innovative design offers distinct advantages over the basic idea, making the system more intelligent and autonomous, thus simplifying maintenance and enhancing result quality.

B. Related Works

To adapt the use of encryption systems to various environmental contexts and types of data exchanged, several research efforts have been undertaken to determine the most suitable encryption approaches. For instance, in [3], the authors directed their investigation toward the Smart Grid (SG), recognizing that devices within this network generate substantial data flows daily. To enhance security in this specific network, [3] proposed an approach grounded in multi-criteria analysis, designed to select the most appropriate encryption algorithm for optimizing energy production, consumption, and distribution. The PROMETHEE method was employed in this work, affirming the effectiveness of the AES-128 algorithm in alignment with decision-makers' preferences. Factors such as memory usage, encryption and decryption times, battery power consumption, and simulation time were taken into account.

Furthermore, the study in [10] represents another notable study that leverages a decision-making approach to determine the lightest and most secure encryption and authentication methods for Internet of Things (IoT) devices, particularly within the realm of IoHT (Internet of Healthcare Things). This research centered on data exchanges among IoHT devices operating within a healthcare environment. These devices generate sensitive data, possess limited processing capabilities, constrained bandwidth, and finite storage memory. The evaluation and decision-making approach introduced in [10] hybridized the CRITIC and TOPSIS methods, considering a diverse array of security criteria in line with the standards set by the International Organization for Standardization (ISO) and the National Institute of Standards and Technology (NIST). The experimentation from this study yielded insightful results, indicating KLEIN cipher as the most lightweight and secure choice among lightweight ciphers, including PRESENT-80, SEA, HIGH, LEA, AES Block Cipher, mCrypton, NOEKEON, Camellia, and the TEA numbers.

The uniqueness of our work lies in the fact that our proposal bases its decision-making on an extensive knowledge base, encompassing a wide array of encryption scenarios and cryptographic algorithms. Thus, our solution transcends the limitations of focusing on a specific domain or a well-defined environment. Instead, it strives to consolidate the wealth of expertise within the encryption field, with the aim of harnessing this knowledge for tailored decision-making in specific cases. Whether it's an ordinary network, an IoT network, or a P2P network, and whether the data is in the form of images, text, videos, or any other format, all these criteria serve as the parameters for our application in selecting the most appropriate encryption algorithm.

II. The Proposed Method

The proposed solution draws from one of the pioneering methods within the realm of artificial intelligence, known for its successful application in various fields. We've harnessed a Knowledge-Based System to address the domain of data encryption.

Our Knowledge-Based System comprises four fundamental modules [12], each encompassing a set of operations for knowledge processing and utilization:

- Module 1: Acquisition of Knowledge
- Module 2: Knowledge Representation
- Module 3: Knowledge Processing
- Module 4: Knowledge Utilization

In general, when crafting a Knowledge-Based System, three vital technical components are essential for system modeling and design [13]:

- The Knowledge Base
- The Fact Base
- The Inference Engine

Incorporating these technical components into a Knowledge-Based System, we adhere to the aforementioned modules, ensuring the system is enriched and leveraged effectively. In this work, we propose to adopt the architectural framework illustrated in the following diagram “Fig. 2” as the foundation for designing our decision-making Knowledge-Based System.

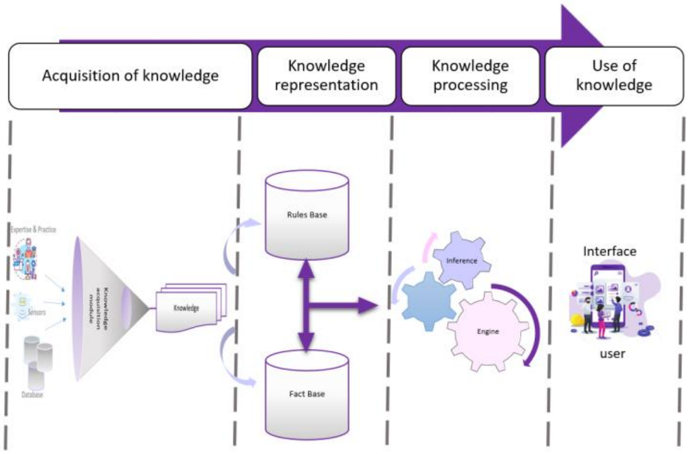


Fig. 2. The KBS Architecture for data ciphering.

This architecture offers precise control of the system, making maintenance more straightforward by segregating data analysis and classification from the decision-making component, which determines the most suitable rule.

To elaborate on the Ciphering Knowledge-Based System, the next section (see Section III) will begin by explaining the underlying principle of our proposed model within the context of data encryption. Subsequently, starting from Section III and continuing through to the end of this section, we will delve into the intricate steps carried out by Modules 1, 2, and 3 within this architectural framework. Notably, Module 4, which represents the graphical user interface (GUI) component of the application, is not covered in the forthcoming sections as it pertains to the user interaction and interface design.

III. Research Method

To construct a system of this nature, it's imperative to first establish a clear definition of knowledge and how it relates to our specific field of application. As articulated in [14], “Knowledge is the mobilization of one or more information in a well-defined context in order to trigger an action or produce knowledge” [14].

In the context of data exchange security, particularly in the pursuit of confidentiality, we will create a knowledge framework that encompasses various criteria influencing this confidentiality. As previously mentioned, these criteria span the source and destination environments, the network, data types, as well as the memory and energy capabilities of the devices involved in the process. Additionally, the encryption system itself must be highly reliable when employed under such conditions. Thus, our system must possess the capability to predict the most appropriate encryption algorithm to ensure that the ciphertext does not reveal any information about the plaintext. This can be quantified through metrics such as entropy or by considering factors like avalanche rate and other critical security criteria.

In the upcoming sections, we will provide a detailed explanation of the parameters that are taken into consideration within this solution.

A. Acquisition of Knowledge

The initial module crucial to the construction of this system is the acquisition of knowledge. Within this module, we must delineate the various potential sources for gathering knowledge pertinent to our issue. This acquisition can be realized through collaboration with experts in the field and/or by interfacing with databases from existing systems, which have been informed by established practices and historical data.

This module unfolds in three pivotal steps:

1) Data classification: As our approach involves data from diverse sources, our initial step is to classify this data based on the environment, structure, and nature of the information [15]. This classification allows us to extract the most critical insights. The outcomes of this classification process are subsequently consolidated in a temporary database.

2) Extraction of information: Next, we move on to extract the most relevant information that directly influences the effectiveness of the encryption system. This includes:

- Message type
- The percentage of images compared to all the data to be encrypted
- The percentage of the video to be encrypted compared to all the data to be encrypted
- The percentage of the Literal text compared to the set of data to be encrypted
- The size of the message
- Device type
- Device capacity
- Network type
- Network size
- Etc.

3) Building knowledge: The objective of this analysis is to enrich the knowledge base with the actual results of the encryption performed.

The goal of this analysis is to enrich the knowledge base with the real-world outcomes of the encryption procedures. During this phase of populating the knowledge base, we lay the foundation for utilizing it in the decision-making process. Here, we execute the integrated encryption algorithms within the system and subsequently compute indicators pertaining to their performance and security levels.

The knowledge we aim to capture during this phase includes:

Encryption execution time

Memory used for encryption

Decryption execution time

Memory used for decryption

Entropy of the encrypted message [16]

This stage is invoked at two distinct points in the life cycle of this approach:

In the Run-in phase: This phase is corresponds to the training phase of the system. It facilitates the generation of new experiences and the enrichment of the decision-making knowledge base.

In the Enrichment phase: This is the final step, occurring after each execution of this method, wherein the selected encryption algorithm is applied.

B. Knowledge Representation

Knowledge can take on various forms, including declarative knowledge, procedural knowledge, and a structured category that combines both, often referred to as meta-knowledge.

Declarative knowledge involves specifying how an action is performed using conditional statements (If... Then...). It's essentially the "facts" within a Knowledge-Based System (KBS). This type of knowledge encapsulates the logic of an operation, defining the objects and concepts that lead to a specific "fact." Declarative knowledge allows for the inclusion of diverse facts from different domains, making it versatile. However, this architecture tends to be slower because it relies on interpreting procedures during each execution.

On the other hand, procedural knowledge already provides instructions, as it outlines the logic of actions in the form of "rules, procedures, strategies, and agendas." The advantage of this architecture is speed, as knowledge is codified in compiled procedures, ready for execution. However, accessing data can be more complex, as it's embedded within the compiled code.

In the realm of AI, knowledge representation in each category is formalized differently and can take on various forms :

In “Fig. 3” we represent an example of the Triplet Knowledge such as: Triplet <object, attribute, value>:

In this context, the "object" refers to the subject that needs to undergo processing, the "attribute" represents the specific property or characteristic of interest, and the "value" corresponds to the specific value associated with this property.. Example:

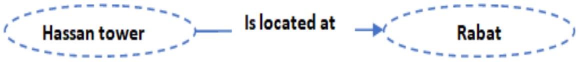
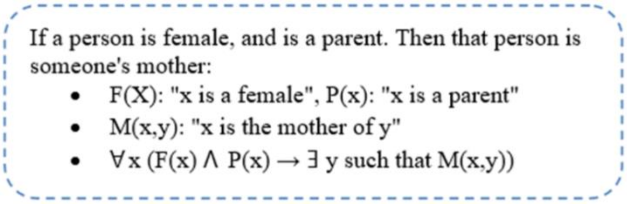


Fig. 3. Example of triplet knowledge representation.

And “Fig. 4” represent Logical formula: Indeed, by employing predicates and propositions in logical formulas, we can effectively depict a particular situation. Example:



If a person is female, and is a parent. Then that person is someone's mother:

- F(X): "x is a female", P(x): "x is a parent"
- M(x,y): "x is the mother of y"
- $\forall x\ (F(x) \land P(x) \rightarrow \exists\ y$ such that $M(x,y))$

Fig. 4. Example of Logical formula of Knowledge representation.

Semantic network: Exactly, in a semantic network, concepts are represented as nodes, and the relationships between these concepts are depicted as arcs or edges. This visual structure “Fig. 5” helps convey how different concepts are interconnected. Example:

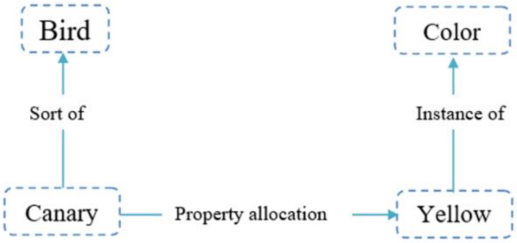


Fig. 5. Example of Semantic network of Knowledge representation

Rule: Establishing connections between pieces of information, thereby extracting additional insights, is pivotal for drawing conclusions regarding relationships, strategies, directives, heuristics, and more. Example:

If <Flower, Color, Rose> Then “I like the Flower”

If “I like the Flower” then “I buy the Flower”

If “I like the Flower” and <Packaging, Price, Free> Then “I buy a bouquet of Flowers”.

The embodiment of knowledge within our Knowledge-Based System (KBS) can be described as a fusion of the Triplet and Rules concepts. Consequently,, to present an encryption experiment applied to certain values of the previously defined criteria, we consider tables for each object. Each table serves as a dimension within our decision analysis framework, with each criterion serving as an attribute within that dimension, encompassing multiple potential values. If we consider the criteria discussed earlier, we will find ourselves working with, at the very least, the following sheme; “Fig. 6”:

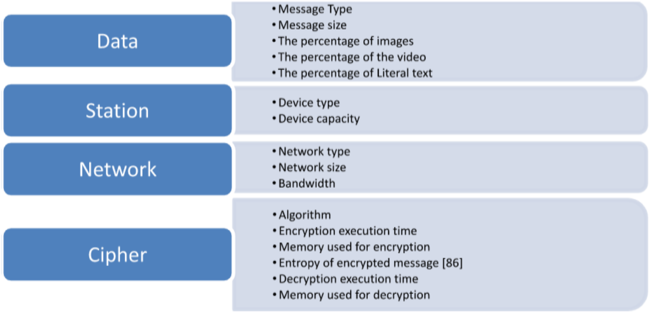

Data
• Message Type
• Message size
• The percentage of images
• The percentage of the video
• The percentage of Literal text
Station
• Device type
• Device capacity
Network
• Network type
• Network size
• Bandwidth
Cipher
• Algorithm
• Encryption execution time
• Memory used for encryption
• Entropy of encrypted message [86]
• Decryption execution time
• Memory used for decryption

Fig. 6. Decision analysis of knowledge-based encryption.

For example, an experience in this knowledge base can be translated as following:

- If (Triplet <Data, type, 'image'> and Triplet <Data, size, '10587'> and Triplet <Station, type, IoT> and Triplet <Cipher, Time, 20ms>) Then Triplet “Cipher, Algo, AES”.

In a general context, each unit of knowledge can be subdivided into two distinct components. The first part embodies the conditions necessary to trigger an action, referred to as "Premises." The second part encompasses the consequences that result from the activation of this action, denoted as "Conclusions." These Conclusions, in turn, correspond to outcomes that may either initiate subsequent actions or determine the ultimate state of affairs, often characterized as "Facts."

To maximize the efficient utilization of the acquired knowledge, we opt for a knowledge base structured around a decision-making architecture in “Fig. 7”. This framework allows for the organization of these various elements into separate, independent structures. Specifically, we establish:

- A database of dimensions: This database consolidates the distinct characteristics pertaining to each object of analysis.

- A database of Facts: Within this repository, we house all the factual information that delineates diverse potential scenarios, along with their corresponding encryption outcomes.

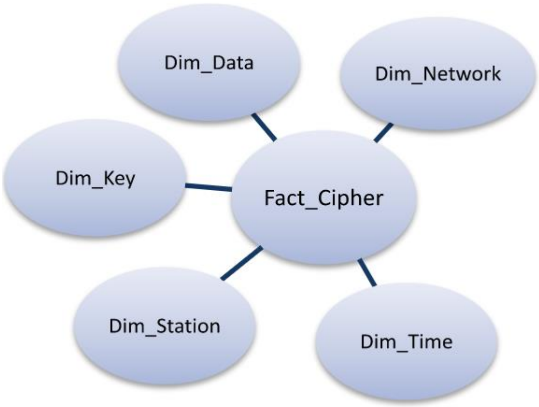


Fig. 7. Decision-making architecture of the knowledge-based encryption.

Going back to the example given above, we have:

- [Triple <Data, type, 'image'> and Triplet <Data, size, '10587'>]: represents a row of the Data dimension with a unique identifier: id_data.
- Triplet <Station, type, IoT>: represents a line of the Station dimension with a unique identifier: id_stat.
- Triplet <Cipher, Time, 20ms>: represents a line of the Cipher dimension with a unique identifier: id_cipher.

The rule is stored in the fact table as follows:

- If (id_data and id_stat and id_cipher) Then id_cipher.

C. *Knowledge Processing*

1) Inference engine: The inference engine, comprising a series of computer instructions, facilitates the process of logical reasoning in alignment with the knowledge and expertise encapsulated within the knowledge base. It harnesses the rule base, executing a sequence of logical inferences and ultimately deducing fresh insights to achieve predefined objectives [17]. To perform its task effectively, the inference engine must be able to detect and handle the following cases:

- Designate the set of rules of the BR to be compared with the facts of the BF.
- Specify the scheduling of these rules in line with the requested need.
- Trigger the execution of the chosen rules according to the sequence strategy previously specified.
- Detect and apply factbase updates.
- Manage duplicate rules and eliminate rules that are already in use.
- Detect rules that can cause confusion and eliminate contradictions.

a) Designate the set of rules: In this phase, we recommend employing the Nearest Neighbor search (NN) algorithm to identify the rules that closely align with our specific real-world case. This algorithm actively queries the knowledge base, taking into account the pre-established criteria, and retrieves all the rows that exhibit dominance concerning these criteria.

b) Nearest Neighbor Search Algorithm (NN)

i) Dominance concept: In a dataset comprising multidimensional objects, a relationship is deemed one of dominance if the dominant object excels in at least one dimension while maintaining a high level of performance across all other dimensions [17]. To identify all such dominant objects within a database of multidimensional entities, we leverage the Skyline operator. This operator empowers our query to retrieve a collection of points that remain unchallenged by any other object, aptly referred to as Skyline points [18].

Illustrating the concept of the Skyline operator, consider the common scenario of seeking a hotel near the beach at a significantly reduced cost. The criteria for this search may often entail trade-offs and appear somewhat contradictory, potentially yielding no results with a conventional selection query. However, it becomes crucial to assist the user in finding a combination that aligns more closely with their preferences. The Skyline operator fulfills this need by presenting the user with a set of the most appealing hotels in accordance with their specified requirements. As depicted in the example below, the curve highlights the Skyline hotels, each of which stands unchallenged by any other hotel in terms of the defined criteria specified in “Fig. 8”.

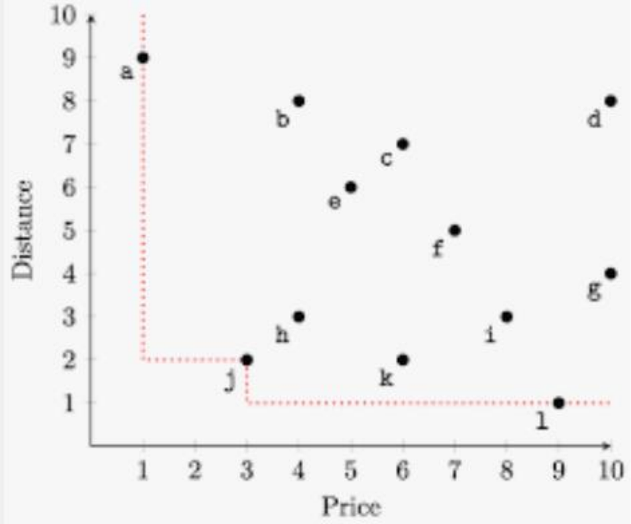


Fig. 8. Curve of elected skyline hotels.

Nonetheless, processing Skyline requests presents a substantial challenge due to the requirement of handling vast amounts of data in memory, a factor that considerably amplifies the algorithmic complexity of this operator [19]. To address this issue, multiple solutions have been developed, focusing on leveraging secondary memory in the Skyline point search process. These solutions fall into two distinct categories: non-index-based algorithms and index-based algorithms.

In our research, we opt for the latter category, specifically embracing index-based algorithms. One such algorithm is the Nearest Neighbor (NN) algorithm, which employs a nearest neighbor search method based on an appropriate distance function suited to various values within the targeted search set. The algorithm effectively partitions the space into regions and systematically identifies points closest to the origin of each region based on a monotonically decreasing distance function, often exemplified by the Euclidean distance. Furthermore, the algorithm continuously evaluates the dominance of candidate objects within each region. Regions dominated by a candidate are progressively eliminated from consideration, and this process continues until the list is ultimately empty [20].

c) Specifying rule ordering: After obtaining all the Skyline points using the Nearest Neighbor (NN) method, the next step involves arranging these points in a user-preferred order and categorizing them based on the name of the encryption algorithm. The chosen encryption algorithm is then identified as the first entry in the sorted list.

d) Detect database updates: The selected encryption algorithm is subsequently applied to the data source to secure its exchange. Simultaneously, the pertinent values associated with the execution of this specific experiment are recorded within our knowledge base.

At this juncture, the application process reaches its conclusion. It's worth noting that the system does not incorporate steps related to managing duplicate rules or addressing potential confusions, as each encryption operation carried out within the system is treated as an independent case.

IV. Results and Discussion

A. Experiment

The tests carried out in this work are based on a technical environment of 16GB of RAM and an Intel(R) Core(TM) i7-6700HQ, 2.60GHz, 64bit OS processor.

The data stored in the knowledge base was generated via an application combining the following five encryption algorithms: AES, Blowfish, DES, TripleDES and ASEC.

The application takes different types of input; it performs an analysis to be able to identify their fixed characteristics (type, size, and environment). Then, it applies the encryption algorithms mentioned above, and as an output, it sends the performance study of each algorithm (execution time, memory capacity used, entropy). Initially, we focused on a sample of 200 different entries, including 100 with the same fixed characteristics.This makes a total of more than a thousand lines of tests.

1) Test scenarios: The experiments conducted within the scope of this contribution were structured around a defined set of test scenarios. Each test case was meticulously designed in accordance with the number of dimensions and their specific types. An effort was made to encompass as many diverse choices as possible, tailored to our specific context. The application, as part of its execution, queries the previously configured knowledge base and deploys the Skyline Nearest Neighbor (NN) algorithm [20], using the input criteria. Subsequently, it returns the Skyline points that correspond to the chosen encryption algorithm.

The primary objective of these experiments is to showcase the efficacy of Skyline algorithms across various specified cases, ranging from scenarios with as few as two dimensions to more complex scenarios with up to ten dimensions. Additionally, some test scenarios incorporate additional dimensions featuring fictitious values, enabling us to consider a broader spectrum of dimensions that are typically associated with network, platform, or transport channel criteria. Unfortunately, not all of these dimensions could be simulated and integrated into the system.

The details of the different test scenarios are outlined in Table I.

TABLE I. Test Scenarios to Experiment the Knowledge-Based Encryption

N°	No. Dimension	Dimensions	Dominance criterion
Scenario 1	2	CipheringRuntime	Min
		Entropy	Max
Scenario 2	2	DecipheringRuntime	Min
		DecipheringMemory	Min
Scenario 3	3	CipheringRuntime	Min
		CipheringMemory	Min
		Entropy	Max
Scenario 4	4	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		Entropy	Max
Scenario 5	6	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
		Entropy	Max
		Dim_fictive_1* (limited No. values)	Min
Scenario 6	6	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
		Dim_fictive_1* (limited No. values)	Max
		Dim_fictive_2* (limited No. values)	Min
Scenario 7	10	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
		Entropy	Max
		Dim_fictive_1 (limited No. values)	Min
		Dim_fictive_2 (limited No. values)	Min

		Dim_fictive_3* (limited No. values)	Min
		Dim_fictive_4* (limited No. values)	Min
		Dim_fictive_5* (limited No. values	Min
Scenario 8	10	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
		Dim_fictive_1 (limited No. values)	Max
		Dim_fictive_2 (limited No. values)	Min
		Dim_fictive_3 (limited No. values)	Min
		Dim_fictive_4 (limited No. values)	Min
		Dim_fictive_5 (limited No. values)	Min
		Dim_fictive_6* (limited No. values)	Min

[* : The Dim_fictive_i represent fictitious dimensions whose values have been generated with an approximate rule to designate other constraints. These dimensions have been added as an indication in order to test the impact of the number of dimensions on the performance of the system.]

B. Results

1) Test of Solution quality: To test the quality of the elected solution, we will focus more on the scenarios with 2, 3 and 4 dimensions containing the entropy and in which we have the real values of their executions. Indeed, focusing on the entropy dimension will allow us to decide whether the chosen cipher is well secured or not since this dimension reflects the amount of information on the source and contained in the cipher.

a) Result of scenario 1: The first scenario goes up 16 Skyline lines, the first four of which all designate the Blowfish algorithm, followed by the AES, ASEC and then 3DES algorithms. The following Table II shows the results of this run:

TABLE II. ELECTED SKYLINE ENCRYPTION OF SCENARIO 1

Algorithm	Runtime	Entropy
BLOWFISH	25	3.8870066417181426
BLOWFISH	49	3.8777830337525954
BLOWFISH	55	3.857533855872884
BLOWFISH	65	3.852401615754532
AES	67	3.821903635277186
AES	73	3.808205259874092
AES	77	3.769139091307226
AES	78	3.7276250338839367
ASEC	107	3.1246923931800934
ASEC	110	2.992846680894221
3DES	123	2.8588450604683873
3DES	229	2.8533695224817235
3DES	334	2.8444354220858403
3DES	353	2.837662090839393
3DES	500	2.830898867431037
3DES	721	2.8282605678748394

b) Result of scenario 2: Scenario 2 pulls up 5 Skyline lines, all of which point to the Blowfish algorithm. The following Table III shows the results of this execution:

TABLE III. ELECTED SKYLINE ENCRYPTION OF SCENARIO 2

Algorithm	Ciphering Time	Memory Used	Entropy
BLOWFISH	25	8.492485106920858	3.8870066417181426
BLOWFISH	29	7.911124302269485	3.9116333559383376
BLOWFISH	37	7.264124532937471	3.903903802424076
BLOWFISH	135	7.234850719105421	3.8894956802766205
BLOWFISH	169	7.231180846946539	3.856438845863682

c) Result of scenario 3: Scenario 3 pulls up 6 Skyline lines, all of which point to the Blowfish algorithm. The following Table IV shows the results of this run:

TABLE IV. ELECTED SKYLINE ENCRYPTION OF SCENARIO 3

Algorithm	Ciph-ering Time	Mem-ory Used	Entropy	Decip-hering Time
BLOWFISH	24	9.09971086299	3.922074380780725	36
BLOWFISH	25	7.7855571424666685	4.003422725072732	37.5
BLOWFISH	39	7.324437668623221	4.028405849999899	58.5
BLOWFISH	94	7.2474787148710185	4.027019099838717	141
BLOWFISH	180	7.236124737916946	3.8650310082953587	270
BLOWFISH	188	7.2223061583614765	3.8935354899007044	282

C. Discussion

1) Solution quality: Based on all the executed experiments, it is evident that the data intended for encryption in this study can be effectively secured using the BLOWFISH algorithm. The system's predictions consistently favored the selection of the BLOWFISH algorithm in all scenarios, particularly those emphasizing the "entropy" dimension, as well as dimensions related to encryption time and memory usage. These dimensions collectively provide a robust basis for evaluating the chosen solution.

Entropy, as a metric, serves as a reliable indicator of the security level and the solution's resistance to various forms of attacks [16], while the dimensions of encryption time and memory usage offer valuable insights into the solution's performance and associated costs.

Furthermore, it's noteworthy that the Skyline points returned by the system exhibit a growing level of agreement, consistently converging on the same solution [18]. This marks a notable departure from the previous encryption method [11], where multiple solutions were viable. Such convergence obviates the need for managing preferences and priorities among the criteria used.

To validate the predictive results and ensure their accuracy, the chosen solution is executed, and pertinent measurements are calculated. The ensuing graph visually depicts the outcome of this comparative analysis.

The following graph shows in “Fig. 9” the result of this comparison:

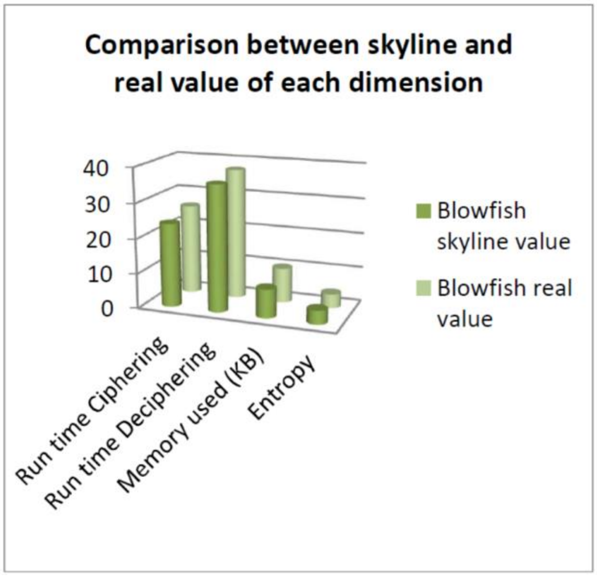


Fig. 9. Comparison between skyline solution value and real value of each dimension.

From this comparison, we can conclude that the difference between the values that were reported by the predictions of the system and the concrete values given by the application of the chosen solution is really negligible. The new values are very close to those predicted by the system.

2) System performance: Now, to see the performance of this system, we ran the set of predefined scenarios and we calculated the time that the system takes to return its results. The graph below “Fig. 10” shows the evolution of the execution time of the system according to the number of dimensions while comparing with the evolution of the old method:

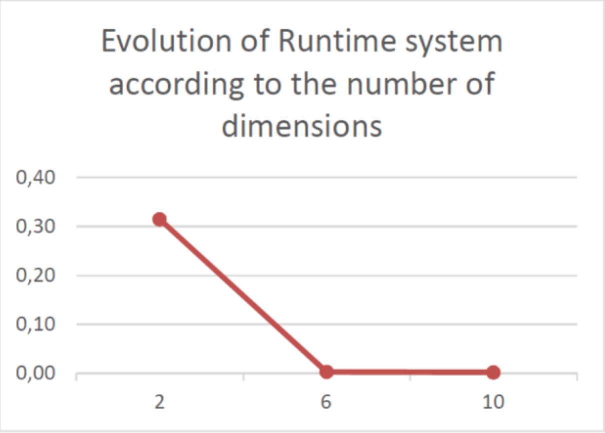


Fig. 10. Evolution of runtime system according to the number of dimensions.

As per the data illustrated in the graph, it becomes evident that the system's performance consistently improves as the number of dimensions increases. This observation suggests that the Nearest Neighbor (NN) algorithm is highly suitable for our configuration, offering enhanced scalability for accommodating various settings. In contrast, the previous intelligent encryption method [11], reliant on the BNL algorithm, experiences a decline in performance as the number of dimensions increases.

The accompanying graph in “Fig. 11” further elucidates this comparative analysis, emphasizing the advantages of the NN algorithm in handling increased complexity and dimensionality, making it a preferred choice for our system.

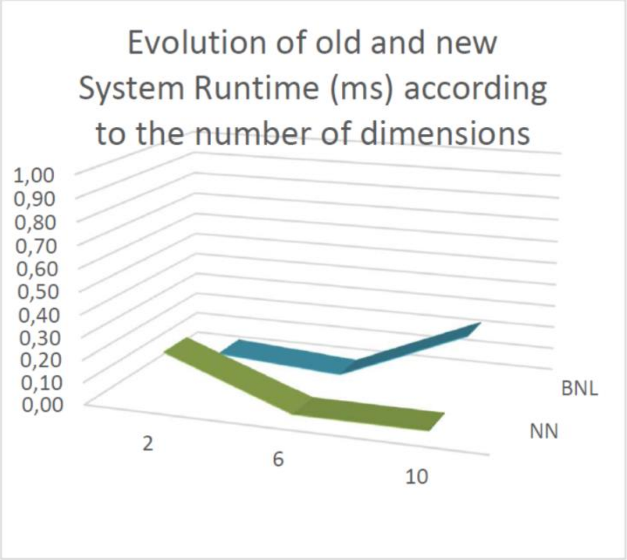


Fig. 11. Evolution of old and new System Runtime (ms) according to the number of dimensions.

3) System maintainability: As previously discussed, the new Knowledge-Based System (KBS) encryption system is constructed upon a modular architecture, where it comprises distinct modules, each responsible for specific functions. This modular design enhances code organization and simplifies system maintenance [22] [23]. Furthermore, the system incorporates a collection of well-established, standardized encryption algorithms, each with a proven track record of robustness and security.

These two key attributes, modularity and standardization, play a pivotal role in the quality of computer system development [21]. They promote system evolution and facilitate maintenance in a flexible and straightforward manner, ultimately contributing to the overall efficiency and effectiveness of the system.

V. Conclusion

Throughout this work, our focus has been on the novel approach to intelligent encryption, one grounded in decisional concepts to enhance the security of data exchanges. To boost its performance, this new approach harnesses the framework of a knowledge-based system. This architectural choice allows us to effectively segregate the processes of data analysis and classification from the construction of knowledge to decision-making. This separation significantly bolsters the system's performance in terms of both the quality of the selected solution and the execution cost.

Our Knowledge-Based System (KBS) for Ciphering incorporates well-defined technical components. The knowledge base, for instance, has been meticulously modeled in a multidimensional fashion, and the inference engine is enriched with the Nearest Neighbor (NN) algorithm within its inference engine to formulate the encryption policy to be adhered to. We have subjected the system to various test cases, each of which explores a different combination by altering the number and type of dimensions (criteria). This approach has allowed us to quantitatively assess the quality and performance of this innovative solution.

Consequently, we've conducted a comprehensive comparative study, juxtaposing the primitives used in this work with the old method. This examination serves to highlight the advantages and advancements brought about by our new architecture, which is built upon separate modules and established security standards. Ultimately, this design choice renders the system more flexible and easier to maintain.

In summary, our system has demonstrated the potential to offer an excellent quality-to-cost ratio for the encryption processes it facilitates, underscoring its efficiency and effectiveness in securing data exchanges.

REFERENCES

[1] Md. M. Ahamad and Md. I. Abdullah, “Comparison of Encryption Algorithms for Multimedia,” *Rajshahi Univ. j. sci. eng.*, vol. 44, pp. 131–139, Nov. 2016, doi: 10.3329/rujse.v44i0.30398.

[2] S. R. Ellis, “Chapter 63 - Fundamentals of Cryptography,” in *Computer and Information Security Handbook (Second Edition)*, J. R. Vacca, Ed. Boston: Morgan Kaufmann, 2013, pp. 1031–1038. doi: 10.1016/B978-0-12-394397-2.00063-5.

[3] R. Mouachi *et al.*, “A Choice of Symmetric Cryptographic Algorithms based on Multi-Criteria Analysis Approach for Securing Smart Grid,” *Indian Journal of Science and Technology*, vol. 10, no. 39, pp. 1–9, Dec. 2017, doi: 10.17485/ijst/2017/v10i39/119856.

[4] F. Omary, A. Mouloudi, A. Tragha, and A. Bellaachia, “A new ciphering method associated with evolutionary algorithm,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3984 LNCS, pp. 346–354, 2006, doi: 10.1007/11751649_38.

[5] S. Trichni, F. Omary, A. Idrissi, M. Bougrine, and M. Abourezq, “New intelligent strategy for encryption decisional support system,” *International Journal of High Performance Systems Architecture*, vol. 9, no. 4, pp. 173–181, 2020, doi: 10.1504/IJHPSA.2020.113678.

[6] M. Bougrine, F. Omary, and S. Trichni, “Security of a new hybrid ciphering system,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, pp. 694–699, 2020.

[7] M. N. A. Wahid, A. Ali, B. Esparham, and M. Marwan, “A Comparison of Cryptographic Algorithms: DES, 3DES, AES, RSA and Blowfish for Guessing Attacks Prevention,” p. 7, 2018.

[8] A. Devi, A. Sharma, and A. Rangra, “Performance analysis of Symmetric Key Algorithms: DES, AES and Blowfish for Image encryption and decryption,” vol. 4, no. 6, p. 6, 2015.

[9] M. Oulehla and D. Malanik, “Comparison of cryptographic methods Triple DES, AES and a method based on the arithmetic of elliptic curves (ECC) on the Android mobile platform. - extended version,” *International Journal of Computers and Communications*, vol. 9, p. 62, Jan. 2015.

[10] L. Ning, Y. Ali, H. Ke, S. Nazir, and Z. Huanli, “A Hybrid MCDM Approach of Selecting Lightweight Cryptographic Cipher Based on ISO and NIST Lightweight Cryptography Security Requirements for Internet of Health Things,” *IEEE Access*, vol. 8, pp. 220165–220187, 2020, doi: 10.1109/ACCESS.2020.3041327.

[11] S. Trichni, F. Omary, and M. Bougrine, “New Smart Encryption Approach based on Multidimensional Analysis Tools,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, pp. 666–675, 2021, doi: 10.14569/IJACSA.2021.0120579.

[12] Systèmes et applications intelligents: Actes de la conférence sur les systèmes intelligents 2021 (IntelliSys) Volume 1.

[13] Janiesch, C., Zschech, P. et Heinrich, K. Apprentissage automatique et apprentissage profond. Marchés électroniques 31.685-695 (2021). https://doi.org/10.1007/s12525-021-00475-2.

[14] C.Zhang, Yu.Xie, H. Bai, B. Yu, W. Li, Y. Gao. “A survey on federated learning” Knowledge-Based Systems,Volume 216, 15 March 2021, 106775

[15] T. Hamon, “Modélisation et Représentation des Connaissances - Introduction,” Institut Galilée - Université Paris 13, 2019.

[16] M. Lefevre, “CM8 : Système à Base de Connaissances,” p. 61, 2020.

[17] N. Sendrier, “Introduction à la théorie de l’information,” p. 70.

[18] J.-L. Ermine, “Les systèmes de connaissances,” p. 145.

[19] M. ABOUREZQ, “Cloud Service Selection using the Skyline and Multi Criteria Decision Aiding,” Mohammed V University of Rabat, 2017.

[20] S. Börzsönyi, D. Kossmann, and K. Stocker, “The Skyline operator,” *Proceedings 17th International Conference on Data Engineering*, 2001, doi: 10.1109/ICDE.2001.914855.

[21] S. Berchtold, C. Böhm, D. Keim, and H. Kriegel, “A cost model for nearest neighbor search in high-dimensional data space,” 1997. doi: 10.1145/263661.263671.

[22] Younoussi, Siham & Roudies, Ounsa. (2016). Capability and maturity model for Reuse: A comparative study. 302-308. 10.1109/CloudTech.2016.7847714.

[23] Zhenan Tu, “Research on the Application of Layered Architecture in Computer Software Development.” Journal of Computing and Electronic Information Management. 11. 34-38. 10.54097/jceim.v11i3.08.

Enhancing Software User Interface Testing Through Few Shot Deep Learning: A Novel Approach for Automated Accuracy and Usability Evaluation

Aris Puji Widodo[1*], Adi Wibowo[2], Kabul Kurniawan[3]
Dept. of Computer Science, Universitas Diponegoro, Semarang, Indonesia[1, 2]
Dept. Information Systems and Operations Management, Vienna University of Economics and Business (WU), Vienna, Austria[3]

Abstract—**Traditional user interface (UI) testing methods in software development are time-consuming and prone to human error, requiring more efficient and accurate approaches. Moreover, deep learning requires extensive data training to develop accurate automated UI software testing. This paper proposes an efficient and accurate method for automating UI software testing using Deep learning with training data limitations. We propose a novel deep learning-based framework suitable for UI element analysis in data-scarce situations, focusing on Few-shot learning. Our framework initiates with several robust feature extraction modules that employ and compare sophisticated encoder models to be adept at capturing complex patterns from a sparse dataset. The methodology employs the Enrico and UI screen mistake datasets, overcoming training data limitations. Utilizing encoder models, including CNN, VGG-16, ResNet-50, MobileNet-V3, and EfficientNet-B1, the EfficientNet-B1 model excelled in the setting of Few-Shot learning with five-shot with an average accuracy of 76.05%. Our proposed model's accuracy was improved and compared to the state-of-the-art method. Our findings demonstrate the effectiveness of few-shot learning in UI screen classification, setting new benchmarks in software testing and usability evaluation, particularly in limited data scenarios.**

Keywords—Deep learning; efficientnet; few-shot; software testing; UI screen classification

I. Introduction

Evaluating a User Interface (UI) in computer science requires expertise and partnership to assess its practical benefits for the intended users [1]. Improved UI design enhances user satisfaction and contributes to increased retention and revenue [2]. Fundamental principles in UI design involve exploring solutions, identifying specific attributes, and actively engaging users in the design process [3]. Despite these principles, the manual testing of UIs remains cumbersome due to the hands-on verification of requirements [4]. Nowadays, computer vision based on neural networks is used to advance UI understanding and address this challenge [5].

This study's main contribution is applying the few-shot learning approach, a notable advancement in computer vision based on neural networks, to effectively classify UI screens. This approach becomes particularly relevant when dealing with limited dataset availability and a diverse range of classes. Our implementation involves applying the few-shot learning technique to the Enrico dataset, a dataset of UI screens, and curating a novel dataset encompassing ten classes that represent common UI mistakes. This strategic combination of few-shot learning and creating a dedicated dataset allows us to explore and enhance UI screen classification in a data-driven manner, especially in scenarios with limited data.

A key advantage of few-shot learning is its ability to handle the constraints of limited sample sizes. This is crucial in fields like UI testing, where obtaining large, diverse datasets can be challenging. Few-shot learning techniques are designed to learn effectively from a few examples, making them ideal for situations where data scarcity is a critical issue. By leveraging this approach, our study aims to demonstrate how advanced computer vision based on neural network (see Fig. 1) techniques can overcome traditional data limitations, opening new avenues for efficient and accurate UI testing.

II. Related Work

Deep learning models have brought numerous benefits to the software development industry, significantly enhancing various tasks, including UI testing [6]. Deep Residual Networks (ResNet) have improved considerably image classification, surpassing previous methods on ImageNet. The successful implementation of ResNet signifies a crucial advancement in utilizing deep learning for image classification tasks [7]. The success of deep learning frameworks extends to automating test case generation and repairing unstable tests in functional UI testing [8].

The application of deep learning in UI understanding, particularly addressing Graphical User Interface (GUI) complexity, has been demonstrated in studies such as OwlEye, which achieved an 85% precision and 84% recall. It uncovers previously-undetected UI display issues in popular Android apps [9]. The use of an enhanced dataset from Rico further illustrates the effectiveness of deep learning in UI topic modeling, achieving notable accuracy in screenshot representation [10]. Challenges in machine learning datasets for UI modeling, mainly due to manual collection limitations, have led to the introduction of large datasets like WebUI, emphasizing the need for enhanced visual UI understanding in domains with limited labeled data [11].

*Corresponding Author.

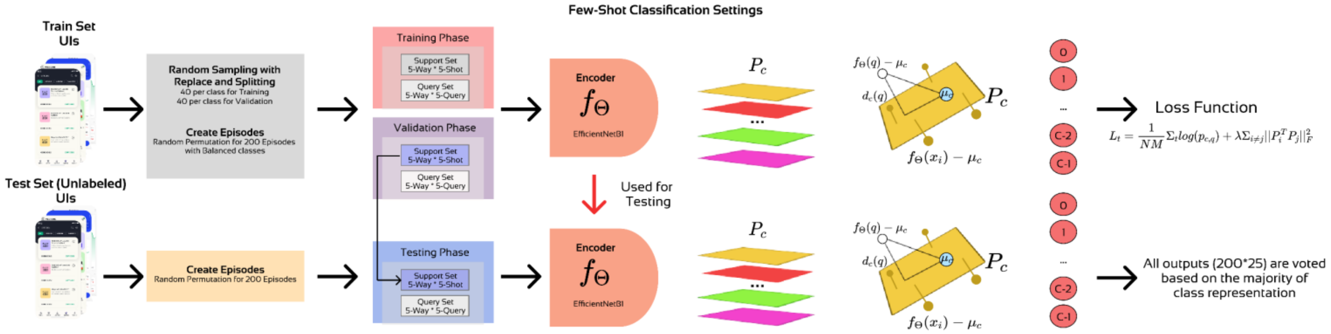

Few-Shot Classification Settings
Train Set UIs
Test Set (Unlabeled) UIs
Random Sampling with Replace and Splitting
40 per class for Training
40 per class for Validation
Create Episodes
Random Permutation for 200 Episodes with Balanced classes
Create Episodes
Random Permutation for 200 Episodes
Training Phase
Support Set 5-Way * 5-Shot
Query Set 5-Way * 5-Query
Validation Phase
Support Set 5-Way * 5-Shot
Query Set 5-Way * 5-Query
Testing Phase
Support Set 5-Way * 5-Shot
Query Set 5-Way * 5-Query
Encoder
f_Θ
EfficientNetB1
Used for Testing
Encoder
f_Θ
EfficientNetB1
P_c
$f_\Theta(q) - \mu_c$
$d_c(q)$
μ_c
P_c
$f_\Theta(x_i) - \mu_c$
0
1
...
C-2
C-1
Loss Function
$L_t = \frac{1}{NM}\Sigma_t log(p_{c,q}) + \lambda\Sigma_{i\neq j}||P_i^T P_j||_F^2$
All outputs (200*25) are voted based on the majority of class representation

Fig. 1. Proposed neural network framework.

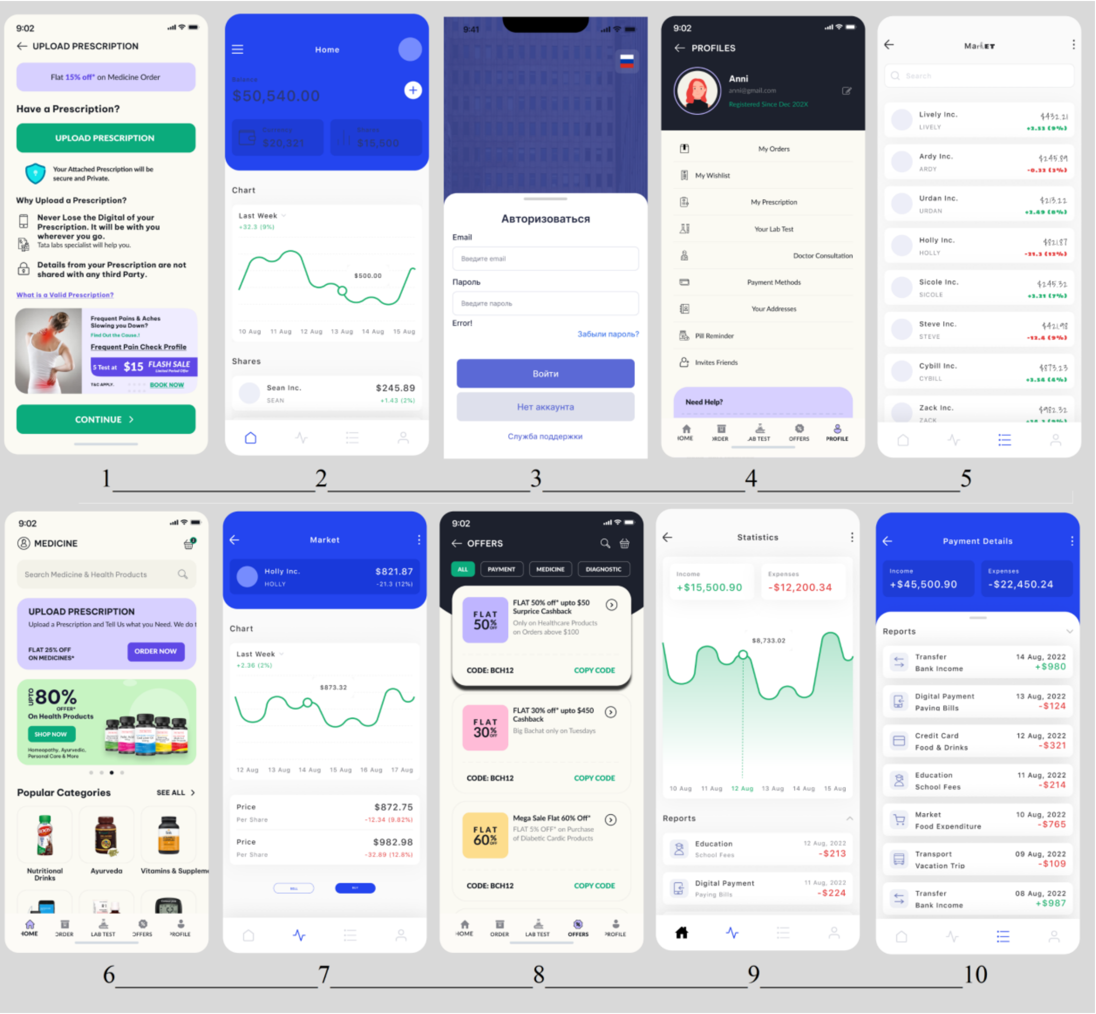


Fig. 2. Sample of the mistake dataset.

Furthermore, researchers thoroughly analyzed current few-shot image classification techniques, categorizing algorithms into transfer learning, meta-learning, data augmentation, and multimodal approaches to address challenges posed by limited sample data [12]. For example, the study introduced GenericConv, a new few-shot learning model for scene classification. Evaluation of benchmark datasets showed that GenericConv successfully addressed imperfections in previous attempts, outperforming benchmark models on three datasets [13]. Another study focused on Few-shot Class Incremental Learning (FSCIL) in real-world applications, introducing the Efficient Prototype Replay and Calibration (EPRC) method. EPRC significantly improved classification performance on CIFAR-100 and miniImageNet compared to mainstream FSCIL methods [14]. In medical imaging, a groundbreaking study proposed a novel few-shot learning method for classifying heart diseases, achieving high segmentation performance and a remarkable 92% accuracy in classifying cardiomyopathy patient groups, even without additional clinical features [15].

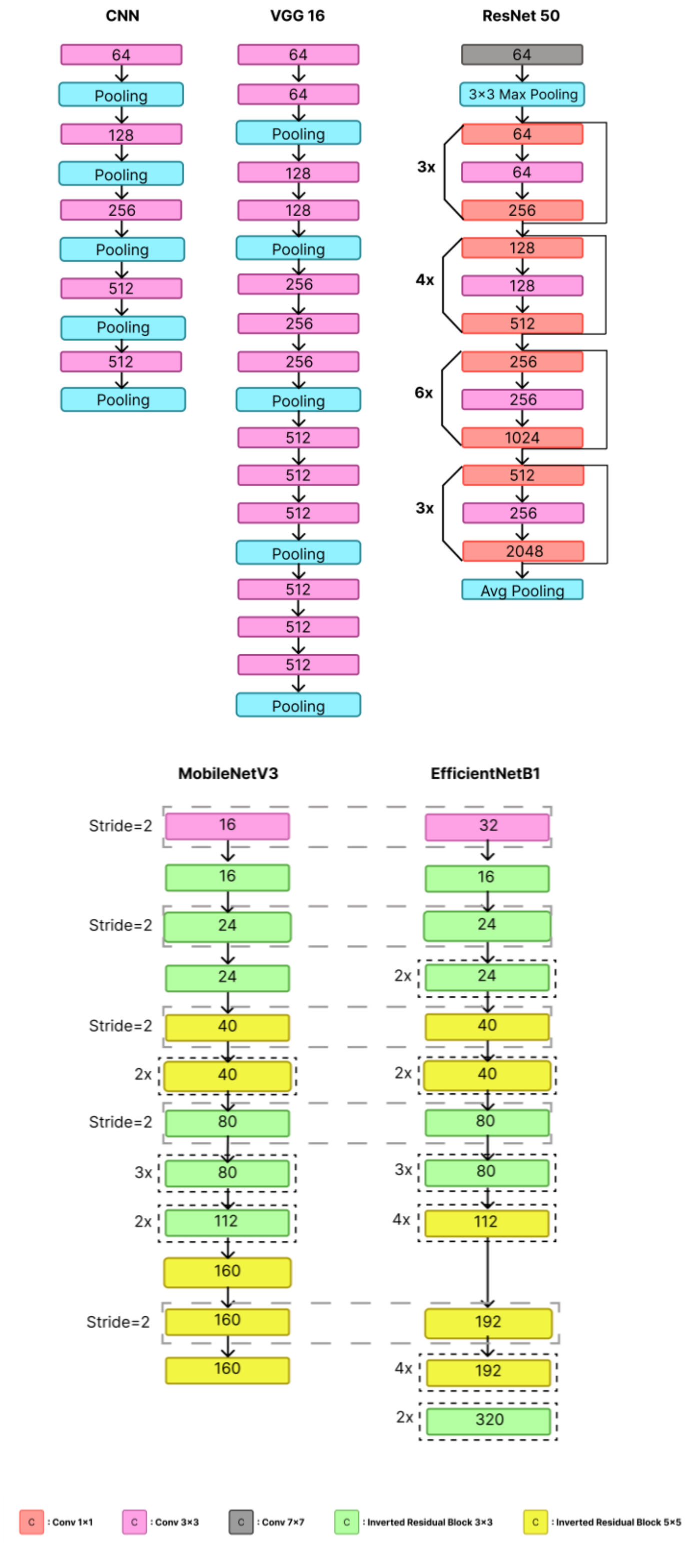


Fig. 3. Encoder Architecture

III. Proposed Method

A. Neural Network Architecture

Our method utilizes a neural network architecture comprising an encoder and an adaptive classifier for few-shot learning. We compare commonly used encoders, including basic CNN, VGG-16, ResNet-50, MobileNet-V3, and EfficientNet-B1, as illustrated in Fig. 3. This experiment marks a novel chapter by leveraging these robust architectures to train the Few-Shot Classification model. The output of these encoders is fed into an adaptive classifier model configured for a few-shot setting.

1) Basic CNN consists of alternating layers of convolution and pooling. Convolutional layers detect local features in the image, while pooling layers reduce the spatial size of the representation, decreasing the number of parameters and computation in the network, which helps prevent overfitting.

2) VGG-16 [16], developed by the Visual Geometry Group at the University of Oxford, is a widely used convolutional neural network (CNN) architecture for image classification. With 16 layers, including 13 convolutional layers and three fully connected layers, VGG-16 excels in feature extraction. The convolutional layers employ small, local filters, capturing hierarchical features as input progresses. The subsequent pooling layers enhance computational efficiency and translation invariance. The final fully connected layers process high-level features, mapping them to specific classes for classification. While VGG-16's deep and intuitive design proves effective in various computer vision tasks, its computational cost limits real-time applications despite being a foundational model for advanced CNN architectures.

3) ResNet-50 [17], developed by Microsoft Research, is a deep convolutional neural network (CNN) renowned for overcoming challenges in training intense networks. Utilizing the concept of residual learning, it introduces skip connections to mitigate the vanishing gradient problem, facilitating the training of deeper networks. Its architecture, featuring residual blocks with skip connections, allows the extraction of intricate features. The initial convolutional layers detect low-level features, while residual blocks, incorporating multiple convolutional layers, use skip connections to learn residual functions. Maximum-pooling layers aid computational efficiency, and fully connected layers map the known features for classification. ResNet-50's innovative architecture enables the training of intense networks, proving highly effective in various computer vision tasks, notably image classification and object recognition.

4) MobileNet-V3 [18], developed by Google, is a lightweight convolutional neural network (CNN) tailored for efficient and accurate computer vision tasks on mobile and edge devices. The model prioritizes high performance while minimizing computational resources, introducing innovative features like inverted residuals and linear bottlenecks to optimize efficiency. Inverted residuals utilize lightweight depthwise separable convolutions, reducing parameters and computational load, while linear bottlenecks balance representational capacity and computational cost. Integration of Squeeze and Excitation (SE) blocks enhances feature recalibration, focusing on crucial channels. MobileNetV3's multiple building blocks collectively form a streamlined and efficient architecture suitable for resource-constrained environments. This design makes it ideal for applications on devices with limited computational resources, offering a commendable trade-off between accuracy and model size in various computer vision tasks, particularly on mobile and edge platforms.

5) EfficientNet-B1 [19], a member of the EfficientNet family developed by Google, is a convolutional neural network (CNN) designed to attain high accuracy with minimal computational complexity. Characterized by balanced scaling of depth, width, and resolution, it optimizes performance across these dimensions. Employing a compound scaling method, EfficientNet-B1 ensures efficient feature extraction at various abstraction levels. Increasing depth, width, and resolution enhances the model's capacity to capture complex and diverse features. Compounding scaling achieves a harmonious balance, maximizing computational resources and performance. This enables EfficientNet-B1 to achieve state-of-the-art accuracy on various computer vision tasks while maintaining a low computational cost. Its adaptability in handling different scaling factors makes it particularly advantageous for resource-constrained environments, striking an optimal balance between accuracy and model size.

B. Few Short Classification

The classification model was built using the few-shot learning approach to deal with restricted data. In itself the meta-learning mode, few-shot terminology is used to train many samples during the training phase; an episode of $\mathcal{T}_i$ was made up of two types of sets: support set $S = \{(x_{1,1}, c_{1,1}), \dots, (x_{N,K}, c_{N,K})\}$ and query set $Q = \{q_1, \dots, q_{N\times M}\}$. The number of S and Q was restricted for each iteration based on N-way, which specifies the number of classes, and K-short or M-query, which denotes the number of samples in each class. The few-shot model is divided into two parts: encoder and classifier.

In this work, a dynamic classifier based on an adaptive subspace [20] was applied. The subspace was used for classification, as well as the shortest distance between the data points and their projection into the subspace. A collection of samples encoded by may be stated as $\tilde{X}_c = [f_\theta(x_{c,1}) - \mu_c, \dots, f_\theta(x_{c,k}) - \mu_c]$, where $\mu_c = \frac{1}{K}\sum_{x_i \in X_c} f_\theta(x_i)$. For instance, q is the query set, and the subspace classifier computation is as follows:

$$d_c = -||(I - M_c)(f_\theta(q) - \mu_c)||^2 \tag{1}$$

where, $M_C = P_C P_C^T$ and μ_c signify the offset between the data point and the subspace, P_C is the truncated matrix of matrix C_C with an orthogonal basis for linear subspace spanning $\mathbb{x}_C = \{f_\theta(x_i); y_i = c\}$ $(hence, B_C^T B_C = I)$.

The chance of a query falling into class c may be calculated using the SoftMax function, which is written as follows:

$$p_{c,q} = p(c|\boldsymbol{q}) = \frac{\exp(d_c(\boldsymbol{q}))}{\sum_{c'} \exp(d_c(\boldsymbol{q}))} \tag{2}$$

Backpropagation through singular value decomposition can be used to minimize the negative log from $p_{c,q}$.

During training, the projection metric on Grassmannian geometry was utilized as a discriminative approach to maximize the margin between two subspaces P_i and P_j, and it is defined as follows:

$$\delta_p^2(P_i, P_j) = \left|\left|P_i P_i^T - P_i P_j^T\right|\right|_F^2 = 2n - 2\left|\left|P_i^T P_j\right|\right|_F^2 \tag{3}$$

the projection metric was maximized by reducing $||P_i^T P_j||_F^2$ and then developing a loss function as stated in Eq. (4); $\mathcal{L}_t$ may then be used to update θ.

$$\mathcal{L}_t = -\frac{1}{NM}\sum_c \log(p_{c,q}) + \lambda \sum_{i\neq j} ||P_i^T P_j||_F^2 \tag{4}$$

IV. Empirical Research Methodology

A. Dataset

1) Enrico dataset [10], is a carefully selected subset originating from Rico, an extensive mobile app dataset. Enrico, a short form for Enhanced Rico, comprises 1,460 UIs categorized into 20 specific design topics. Each category delineates particular UI features, contributing to a detailed comprehension of mobile app design. These topics encompass diverse aspects such as Bare (largely unused area), Dialer (number entry), Camera (camera functionality), Chat (chat functionality), Editor (text/image editing), Form (form-filling functionality), Gallery (grid-like layout with images), List (elements organized in a column), Login (input fields for logging), Maps (geographic display), MediaPlayer (music or video player), Menu (items listed in an overlay or aside), Modal (popup-like window), News (snippets list: image, title, text), Other (everything else, considered as a rejection class), Profile (information on a user profile or product), Search (search engine functionality), Settings (controls to change app settings), Terms (terms and conditions of service), and Tutorial (onboarding screen).

2) The Mistake Dataset is a dataset proposed by this paper as shown in Fig. 2. This dataset consists of 200 user interfaces (UI) categorized into 10 specific design topics. Each category describes specific UI features that contribute to a detailed understanding of mobile device UI design. These topics encompass various aspects such as pointless inconsistency design, inapproriate use of shadows, lack of text hierarchy, bad iconnography, unaligned elements, low contrast, poor typography choices, tiny touch targets, text overlap, and error message clarity. For detailed explanations regarding the 10 design topics, (see Table I).

TABLE I. DESCRIPTION MISTAKE DATASET

No	UI Component	Description
1	Pointless inconsistency design	Identifying inconsistencies in layout across different screens or resolutions
2	Inapproriate use of shadows	Pay attention to the depth of the shadow, they should create a sense of realism and hierarchy
3	Lack of text hierarchy	The text format must have a contract between each text style and the title style
4	Bad iconnography	Ensuring icons are correctly displayed and recognizable
5	Unaligned elements	Detecting misaligned text, buttons, image or other
6	Low contrast	Identifying poor contrast between text and background, affecting readability
7	Poor typography choices	Font selection, font size, spacing and alignment must be taken into account to facilitate readability
8	Tiny touch targets	The size of the touch target must be taken into account with the size of the screen
9	Text overlap	Spotting instances where text overlaps with UI elements
10	Error message clarity	Ensuring that error message are clear, concise, and appropriately placed

B. UI Understanding and Methodology

The initial technologies developed for understanding app user interfaces operate at either the application or screen level, aiming to inspire new designs by exploring existing relevant ones. For instance, [21] creates a searchable gallery of UI element ideas based on app pictures. Enrico [10] draws inspiration from extensive datasets that are too vast to browse effectively, showcasing the evolving landscape of UI exploration.

To enhance design analysis and automation, screen type or functionality classification from a screenshot proves beneficial. Enrico [10], a dataset with 1460 samples (a subset of Rico [22]), categorizes each screenshot into one of 20 design categories. However, the limited dataset size poses challenges for training deep learning classification models. Despite having a vast online dataset, it lacks screen type annotations, preventing the utilization of the pre-training method employed for element recognition.

In the methods section of our research, we present a strategic solution to overcome the limitations of small datasets in training deep learning models for UI screen classification. By integrating a Few-Shot Learning Approach for UI Screen Classification, we enable our models to perform effectively with the Enrico [10] dataset, which comprises a modest collection of 1460 samples. This approach is tailored to efficiently learn from a constrained number of data points, thus allowing accurate classification of each screenshot into one of the proposed 20 design categories. The few-shot learning technique is pivotal in our methodology, as it addresses the challenge of dataset scarcity, a common obstacle in the realm of UI design analysis.

Further augmenting our methodological framework, we have developed a dataset named “Mistake of UI Screen”, encompassing 10 classes of prevalent UI design errors, alongside the use of the Enrico dataset. This dataset is crafted to refine our models' ability to not only discern various UI elements but also to detect and classify typical design flaws systematically. The development of this dataset, alongside the application of few-shot learning, constitutes a robust approach to enhancing the precision of screen classification, thereby contributing significantly to the field of UI design and evaluation.

V. EXPERIMENTAL RESULT

In this study, a few-shot learning approach was applied using Enrico dataset with 20 class and Mistake dataset with 10 class. Randomly selects 40 data per class and splits them to 50% training and 50% validation phase with balanced class distribution. During testing phase, we randomly sampled query images along with the support set from the validation phase, repeating this process twenty-five times and ultimately applying majority voting in a few-shot setting. Settings used were five-way for all phases, five-shot on support and query sets, and 200 episodes for each epoch. Training iteration using 10 epochs for Enrico dataset referring to original paper and 25 epochs for Mistake dataset with learning rate of 1 x 10^{-3} optimized by Adam and lambda of 0.03. Few-shot model ran on an i7 processor with RTX 2060 SUPER.

A. Model Implementation

1) Model and shot setting in enrico dataset: Our assessment of model performance involved a sequence of experiments using the validation subset of the Enrico dataset. Table II compiles the accuracy figures for five models: CNN, VGG-16, ResNet-50, MobileNet-V3, and EfficientNet-B1, across five separate experiments. The table outlines each model's individual experiment results, along with their respective average accuracies and standard deviations. EfficientNet-B1 emerged as the top performer, boasting an average accuracy of 76.05% with a standard deviation of 1.1618%. The findings suggest that models, particularly those with depth and designed for efficiency that leverage pre-trained networks, have the potential to enhance generalizability.

The comparison across models were proposed to further investigated the influence of training variation on the performance of EfficientNet-B1. Table III shows the accuracy results for EfficientNet-B1 when trained with varying shots numbers on validation subset of the Enrico dataset. These conditions ranged from 1 to 5-shot training, with each tested over five experimental runs. The data indicates that increasing the number of shots enhances both the accuracy and stability of the model's performance. Definitely, the five-shot trained EfficientNet-B1 achieved a highest average accuracy, prominence that a greater number of shots correlate with improved model performance.

TABLE II. ACCURACY RESULT FOR DIFFERENT MODELS APPLIED TO THE ENRICO DATASET ON THE VALIDATION SUBSET

Model	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5	Average	Standard Deviation
CNN	0.4020	0.4219	0.4108	0.3976	0.4008	0.4066	0.009843
VGG-16	0.3120	0.4020	0.3506	0.3013	0.3702	0.3472	0.041501
ResNet-50	0.4800	0.4480	0.4373	0.5164	0.5004	0.4764	0.033632
MobileNet-V3	0.7012	0.7230	0.7012	0.7148	0.7119	0.7104	0.009349
EfficientNet-B1	0.7650	0.7432	0.7506	0.7703	0.7732	0.7605	0.011618

TABLE III. ACCURACY RESULT FOR EFFICIENTNET-B1 WITH VARIOUS SHOTS APPLIED TO THE ENRICO DATASET ON THE VALIDATION SET

Number of Shots	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5	Average	Standard Deviation
1-shot	0.5078	0.6800	0.5509	0.5509	0.5860	0.5751	0.064853
2-shot	0.6800	0.6530	0.6230	0.6002	0.6230	0.6358	0.031002
3-shot	0.6800	0.7100	0.6980	0.6800	0.7098	0.6956	0.015012
4-shot	0.7011	0.7130	0.7266	0.6980	0.7100	0.7097	0.011263
5-shot	0.7650	0.7432	0.7506	0.7703	0.7732	0.7605	0.011618

TABLE IV. ACCURACY RESULT FOR EFFICIENTNET-B1 WITH 5- SHOT APPLIED TO THE MISTAKE DATASET ON THE VALIDATION SET

Experiment	Accuracy
Experiment 1	0.3737
Experiment 2	0.4800
Experiment 3	0.3820
Experiment 4	0.4670
Experiment 5	0.4302
Average	0.4266
Standard Deviation	0.0431

2) *Model and shot setting in mistake dataset:* The assessment of the performance EfficientNet-B1 model with five-shot by testing were applied on the validation subset of the Mistake Dataset. The results of this investigation are shows in Table IV about details the model's accuracy across five experiments. These experiments were designed to evaluate the model's ability to predict accurately under different conditions. The overall performance is summarized by the mean accuracy, calculated to be 42.66%. Additionally, the standard deviation of the accuracy, at 0.0431, indicates a relatively small variability in the model's performance throughout the trials, suggesting a stable accuracy profile for the EfficientNet-B1 on this dataset.

B. Comparison with other Method

Table V shows a brief accuracy comparison of various models on Enrico dataset. It contrasts the performance of a model trained on the Screenshot data from the Enrico dataset, achieving 75.8% accuracy, with that of the VGG-16 and Noisy ResNet-50 models trained on the Enrico dataset, which have accuracies of 47.4% and 46.5% respectively. Our adaptation of the EfficientNet-B1 model also performed notably well, with an accuracy close to the top-performing Screenshot model at 76.1%. These outcomes demonstrate the strong performance of our EfficientNet-B1 model, which nearly matches the leading model and substantially exceeds the performance of established architectures like VGG-16 and Noisy ResNet-50. EfficientNet-B1 achieves higher accuracy than CNN, VGG-16, ResNet-50, and MobileNet-V3 due to its optimized balance of depth, width, and resolution in the network architecture. It is designed using a compound scaling method to scale these three dimensions efficiently and in harmony. This results in more effective and efficient use of computational resources and better performance on limited data, as in Few Shot learning scenarios, making it particularly suitable for the complex task of UI screen classification.

TABLE V. ACCURACY COMPARISON ON OTHER METHOD

Model Configuration	Accuracy
Screenshot [10]	75.8%
VGG-16 [23]	47.4%
Noisy ResNet-50 [23]	46.5%
EfficientNet-B1 (our)	**76.1%**

VI. DISCUSSION

In the discussion section of our analysis, we delve into the substances and insights derived from our research findings. We initiate by examining our model's impact on the Enrico and Mistake datasets, scrutinizing how the few-shot learning approach adapts to each dataset's unique characteristics. The transition was used to discuss the broader implications of our model on UI screen classification, highlighting the advancements and the potential it unlocks for future applications. Finally, we address the limitations encountered during our research and offer suggestions for future studies to build upon our work.

A. Model Impact on each Dataset

The application of our neural network model to the Enrico dataset demonstrated remarkable adaptability, as evidenced by the high accuracy rates achieved across diverse UI screen types. This success is attributed to the model's ability to learn from limited examples, a testament to the efficacy of few-shot learning. Conversely, when applied to the Mistake dataset, which contained a different array of UI screen errors, the model faced distinct challenges, reflecting the nuances and complexity of classifying a wider variety of mistakes. These observations underscore the need for tailored approaches when dealing with datasets of varying natures.

B. Improved UI Screen Classification

Our model represents a significant step forward in the realm of UI screen classification. By leveraging few-shot learning, we have shown that it is possible to achieve high levels of precision with limited training data, a scenario common in real-world settings. This advancement is particularly promising for the development sector, where rapid and accurate UI assessment can streamline the design process, reduce costs, and enhance the end-user experience.

C. Limitation and Suggestion

Our primary constraint was the size and diversity of the datasets, which may affect the model's ability to generalize across a broader range of UI designs. For future work, we suggest expanding the datasets to include a more varied set of UI screens and errors. Further, investigating other few-shot learning configurations and incorporating user feedback in the training loop could refine the model's performance and applicability.

VII. Conclusion

In this paper, we propose an efficient and accurate method for automating user interface (UI) testing using Deep learning with training data limitations. This research proposes a novel deep learning-based framework that marks a pivotal advancement in user interface (UI) testing, demonstrating the powerful capabilities of few-shot learning in UI screen classification. The framework initiates with several robust feature extraction modules that employ and compare sophisticated encoder models (CNN, VGG-16, ResNet-50, MobileNet-V3, and EfficientNet-B1) to be adept at capturing complex patterns from a sparse dataset. EfficientNet-B1 achieves higher accuracy than CNN, VGG-16, ResNet-50, and MobileNet-V3 due to its optimized balance of depth, width, and resolution in the network architecture. It is designed using a compound scaling method to scale these three dimensions efficiently and in harmony. This results in more effective and efficient use of computational resources and better performance on limited data, as in Few Shots learning scenarios, making it particularly suitable for the complex task of UI screen classification. Moreover, our proposed model's accuracy was improved and compared to the state-of-the-art method using the Enrico dataset. Our utilization of the Enrico and Mistake datasets confirmed the model's ability to accurately classify a broad spectrum of UI screens with limited training data, illuminating its proficiency in detecting and categorizing intricate UI errors. For the rapidly evolving domain of software development, our model offers a scalable and efficient solution to the perennial challenge of UI testing. This is crucial for the development sector, where quick, reliable UI assessments can significantly expedite the design process, cut costs, and elevate the end-user experience. Moreover, this study paves the way for future research endeavors. The challenges we encountered due to our datasets limited size and diversity underline the need for more expansive and varied data in further investigations. Such efforts could refine the model's applicability and accuracy, integrating user feedback and diverse learning configurations into the development process.

Acknowledgment

This work was supported by the Research for International Publication (RKI) NonAPBN Universitas Diponegoro Indonesia under Grant 609-78/UN7.D2/PP/VIII/2023.

References

[1] N. Samrgandi, “User Interface Design & Evaluation of Mobile Applications User Interface Design & Evaluation of Mobile Applications,” no. February, 2021.

[2] K. Edson et al., “An Evaluation Framework for User Experience Using Eye Tracking , Mouse Tracking , Keyboard Input , and Artificial Intelligence : A Case Study,” International Journal of Human–Computer Interaction, vol. 00, no. 00, pp. 1–15, 2021, doi: 10.1080/10447318.2021.1960092.

[3] M. Bakaev, M. Speicher, J. Jagow, S. Heil, and M. Gaedke, “We Don’t Need No Real Users?! Surveying the Adoption of User-less Automation Tools by UI Design Practitioners,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13362 LNCS, no. July, pp. 406–414, 2022, doi: 10.1007/978-3-031-09917-5_28.

[4] Z. Khaliq, D. A. Khan, and S. U. Farooq, “Using deep learning for selenium web UI functional tests: A case-study with e-commerce applications,” Engineering Applications of Artificial Intelligence, vol. 117, no. August 2022, p. 105446, 2023, doi: 10.1016/j.engappai.2022.105446.

[5] G. Ang and E. P. Lim, “Learning User Interface Semantics from Heterogeneous Networks with Multimodal and Positional Attributes,” International Conference on Intelligent User Interfaces, Proceedings IUI, pp. 433–446, 2022, doi: 10.1145/3490099.3511143.

[6] F. H. Alshammari, “Trends in Intelligent and AI-Based Software Engineering Processes: A Deep Learning-Based Software Process Model Recommendation Method,” Computational Intelligence and Neuroscience, vol. 2022, 2022, doi: 10.1155/2022/1960684.

[7] M. Shafiq and Z. Gu, “Deep Residual Learning for Image Recognition: A Survey,” Applied Sciences (Switzerland), vol. 12, no. 18, pp. 1–43, 2022, doi: 10.3390/app12188972.

[8] Z. Khaliq, S. U. Farooq, and D. A. Khan, “A deep learning-based automated framework for functional User Interface testing,” Information and Software Technology, vol. 150, no. December 2021, p. 106969, 2022, doi: 10.1016/j.infsof.2022.106969.

[9] Z. Liu, C. Chen, J. Wang, Y. Huang, J. Hu, and Q. Wang, “Owl Eyes: Spotting UI Display Issues via Visual Understanding,” Proceedings - 2020 35th IEEE/ACM International Conference on Automated Software Engineering, ASE 2020, pp. 398–409, 2020, doi: 10.1145/3324884.3416547.

[10] L. A. Leiva, A. Hota, and A. Oulasvirta, “Enrico: A Dataset for Topic Modeling of Mobile UI Designs,” Extended Abstracts - 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services: Expanding the Horizon of Mobile Interaction, MobileHCI 2020, 2020, doi: 10.1145/3406324.3410710.

[11] J. Wu, S. Wang, S. Shen, Y. H. Peng, J. Nichols, and J. P. Bigham, “WebUI: A Dataset for Enhancing Visual UI Understanding with Web Semantics,” Conference on Human Factors in Computing Systems - Proceedings, 2023, doi: 10.1145/3544548.3581158.

[12] Y. Liu, H. Zhang, W. Zhang, G. Lu, Q. Tian, and N. Ling, “Few-Shot Image Classification : Current Status and Research Trends,” 2022.

[13] M. Soudy and Y. M. Afify, “GenericConv : A Generic Model for Image Scene Classification Using Few-Shot Learning,” pp. 1–13, 2022.

[14] W. Zhang and X. Gu, "Few Shot Class Incremental Learning via Efficient Prototype," 2023.

[15] A. Wibowo et al., "Cardiac Disease Classification Using Two-Dimensional Thickness and Few-Shot Learning Based on Magnetic Resonance Imaging Image Segmentation," Journal of Imaging, vol. 8, no. 7, 2022, doi: 10.3390/jimaging8070194.

[16] M. Ye et al., "A Lightweight Model of VGG-16 for Remote Sensing Image Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 6916–6922, 2021, doi: 10.1109/JSTARS.2021.3090085.

[17] B. Koonce, "ResNet 50," Convolutional Neural Networks with Swift for Tensorflow, pp. 63–72, 2021, doi: 10.1007/978-1-4842-6168-2_6.

[18] J. Huang et al., "BM-Net: CNN-Based MobileNet-V3 and Bilinear Structure for Breast Cancer Detection in Whole Slide Images," Bioengineering, vol. 9, no. 6, 2022, doi: 10.3390/bioengineering9060261.

[19] C. Wang and Y. Li, "Motion Prediction for Autonomous Vehicles Based on EfficientNet-B1," 2022 4th International Conference on Communications, Information System and Computer Engineering, CISCE 2022, pp. 648–651, 2022, doi: 10.1109/CISCE55963.2022.9851007.

[20] C. Simon, P. Koniusz, R. Nock, and M. Harandi, "Adaptive subspaces for few-shot learning," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 4135–4144, 2020, doi: 10.1109/CVPR42600.2020.00419.

[21] C. Chen, S. Feng, Z. Xing, L. Liu, S. Zhao, and J. Wang, "Gallery D.C.: Design search and knowledge discovery through auto-created GUI component gallery," Proceedings of the ACM on Human-Computer Interaction, vol. 3, no. CSCW, 2019, doi: 10.1145/3359282.

[22] B. Deka et al., "Rico: A mobile app dataset for building data-driven design applications," UIST 2017 - Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 845–854, 2017, doi: 10.1145/3126594.3126651.

[23] J. Wu, S. Wang, S. Shen, Y. H. Peng, J. Nichols, and J. P. Bigham, WebUI: A Dataset for Enhancing Visual UI Understanding with Web Semantics, vol. 1, no. 1. Association for Computing Machinery, 2023.

Optimizing Crop Yield Prediction in Precision Agriculture with Hyperspectral Imaging-Unmixing and Deep Learning

Dr.Deeba K[1*], Dr O.Rama Devi[2], Dr. Mohammed Saleh Al Ansari[3], Dr.Bhargavi Peddi Reddy[4],
Dr. Manohara H T[5], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[6], Manikandan Rengarajan[7]
Assistant Professor, School of Computer Science and Applications, REVA University, Bangalore[1]
Professor and HOD, Department of AI and DS, Lakireddy Balireddy College of Engineering, Mylavaram, Vijayawada[2]
Associate Professor, College of Engineering-Department of Chemical Engineering, University of Bahrain, Bahrain[3]
Associate Professor, Dept of CSE, Vasavi College of Engineering, Hyderabad, India[4]
Assistant Professor, Department of Electronics and Communication Engineering,
NITTE Meenakshi Institute of Technology, Bengaluru[5]
Faculty of Informatics and Computing, UniSZA University, Malaysia[6]
Vel Tech Rangarajan Dr. Sagunthala R and D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India-600062[7]

***Abstract*—The optimization of crop yield projections has arisen as a major problem in modern agriculture, due to the increasing demand for food supply and the necessity for effective resource management. Precision and scalability are hampered by the limits associated with conventional agricultural production prediction techniques, which mostly rely on observations and simple data sources. While methods like random forest (RF) and K-nearest neighbors (KNN) are widely used, their reliance on personal assessments and insufficient knowledge of crop attributes typically results in less accurate forecasts and makes them unsuitable for agricultural precision. The suggested method combines deep learning, spectral unmixing, and hyperspectral imaging methods to overcome these obstacles. With the use of hyperspectral imaging, which records a vast array of data that is not visible to the human eye, crop attributes may be thoroughly examined and can identify the unique spectral fingerprints of different agricultural constituents by using spectral unmixing approaches, which makes it easier to evaluate the health and growth phases of the crop. Then, using this augmented spectral data, deep learning algorithms create a solid, data-driven basis for precise crop production prediction. MATLAB has been used in the suggested workflow. The combination of deep learning, spectrum unmixing, and hyperspectral imaging provides a comprehensive, cutting-edge approach that goes beyond the constraints of conventional techniques were implemented in python. Some of the algorithms that were examined, this one with integration has the lowest Root Mean Square Error (RMSE) of 0.15 and Mean Absolute Error (MAE) of 0.14, demonstrating higher prediction accuracy above other current models. This novel method represents a substantial breakthrough in precision agriculture while also improving crop production prediction.**

Keywords—Crop yield prediction; hyper spectral image; spectral unmixing; resource management; precision agriculture

I. Introduction

Demand for premium agricultural products will increase exponentially as people's standard of living rise. The amount of farmland has, regrettably, decreased due to environmental harm. Therefore, to meet increasing need for food, livestock and agricultural producing operations have become more importance. With the goal of reducing the financial and environmental costs associated with food production, precision agriculture is a technique for boosting productivity [1]. Crop conditions are measured using remote sensing, which is subject to large variation. In order to manage resources and make judgments on crop development, agriculturalists need to have equipment that is technologically advanced. By giving information on crop health and development phases, hyperspectral photography facilitates targeted farming by allowing for effective insect and herbicide treatments. Globally, there is a growing need for modern technology, namely multispectral and hyperspectral pictures, to increase farming precision and control [2]. The hyperspectral images are able to distinguish between artefacts and physicals in a wide range of application fields, including precise agriculture, minerals recognition, analysis of the environment, and urban development [3].

Agriculture and forestry are anticipated to hold the biggest market share among other end-user sectors in the hyperspectral imagery industries. Hyperspectral imaging is utilized in farming and forestry for a number of tasks, including weed mapping, plant recognition, seed yield analysis, and forest management [4]. In addition, over the past ten years, sensors have collected an increasing amount of data on farms. Therefore, offers for data optimization as well as apps for farmers have been coordinated by hyperspectral service providers [5]. Key elements of successful agriculture include the monitoring of nutrient crops, water stress, disease, pest infarction, and general plant health. By using conventional optical detection methods like imaging or spectroscopy, it is difficult to guarantee adequate spatial and spectrum data for the analysis of food and agriculture harvests [6]. The limitations of conventional imaging approaches for sorting vegetables and fruit include their inability to separate internal from exterior product structures and to collect spectral information. The commonly used systems imitate human vision using colour video cameras; nevertheless, these approaches are costly, time-consuming, and frequently result in sample obliteration. Furthermore time-consuming, costly, and sometimes obliterative, current procedures make it challenging to find product flaw [7].

Due to the rapid development of information science, image analysis, and precision agriculture over the past few years, optical detectors have developed into scientific tools. Particularly, the incorporation of a strategy to produce a spectral variation spatial map has led to substantial study and development in imaging and spectroscopic techniques, which has contributed to several well-liked applications in agriculture [8]. Precision agricultural methods have expanded and been more widely used as a result of the development of airborne and ground-based hyperspectral and multispectral imagery equipment. Along with its predictive skills, this technology has made it possible to characterize soils and vegetative cover, evaluate crop pressures, identify bruises in fruits and vegetables and estimate yield [9]. Hyperspectral and multispectral images provide a number of benefits, including inexpensive prices (in comparison to traditional scouting), reliable, simple use, rapid, non-destructive, extremely precise assessments, and a wide range of usages. Typical spectral images are made up of a number of monochromatic images that represent various wavelengths [10]. In comparison to traditional computer vision as well as human perception, hyperspectral imaging systems offer a natural advantage. Using hyperspectral imaging systems, any appearance features that are challenging or difficult to extract with systems may be retrieved [11]. A significant use for hyperspectral imaging is the assessment of the overall quality of agricultural and food items.

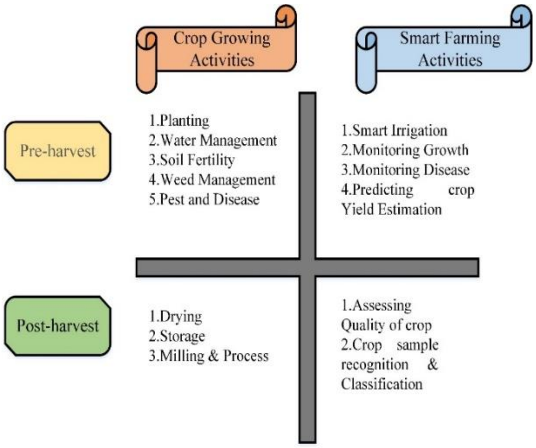


Fig. 1. Process in agriculture.

Fig. 1 shows pre-harvest and post-harvest stages of crop-growing and smart farming practices. The first step is planting, which is followed by direct sowing or transplanting. Water levels are maintained via smart irrigation according to development stages. To guarantee a sufficient supply of water and lessen the load of field weeds, weed control is essential. This strategy lowers the cost of agriculture [12]. The monitoring of soil fertility is a crucial for maximizing plant development. In order to minimize losses, prompt and accurate identification of crop diseases and pest management are crucial. For example, categorizing, identifying, and forecasting infestation patterns in fields are examples of operations included in the agricultural disease monitoring.

Analyzing crop growth utilizing vegetation, remotely sensed data, and climate variables is known as crop growth monitoring. There is also mapping of crop-growing zones at the field level. Vegetation indices, remotely sensed data, and hyper spectral data are used to estimate agricultural production. Harvesting, managing, organizing, cleaning, and transporting are examples of post-production jobs. The quality of crops may be assessed using machine learning techniques [13]. Activities like evaluating the crop's quality or looking into how climate change may affect crop production are frequently included in determining the crop's quality. The crop will next be dried using conventional or mechanical methods in the following stage of the process. The milling procedure, which eliminates the husk, is the final step in the post-production stage. Using image processing and machine learning methods, classification is to distinguish and categorize crop sample objects based on color and texture properties. Key contributions of this research are,

- Advanced Sensing Technology Integration: The research introduces a pioneering approach by utilizing Unmanned Aerial Vehicles (UAVs) equipped with hyper spectral sensors, showcasing the integration of cutting-edge sensing technologies for detailed spectral data collection in agricultural landscapes.

- Precise Hyper spectral Data Processing: The study emphasizes meticulous data pre-processing techniques, including radiometric calibration and geometric corrections using the Hyperspace program. This ensures the accurate conversion of digital data into radiance data, enabling the separation of mixed spectral signals associated with crops, soil types, and other agricultural variables.

- Innovative CNN Framework for Feature Extraction: The research employs a Convolutional Neural Network (CNN) framework to extract key features from hyper spectral data, such as NDVI, CCCI, CVI, contrast, and entropy indices. This innovative approach enhances insights into crop health and landscape dynamics, contributing to the field of precision agriculture.

- Optimized CNN-LSTM Model for Crop Yield Prediction: The study introduces a novel Optimized CNN-LSTM model for accurate crop yield prediction. By integrating deep learning with Firefly Algorithm optimization, the model leverages hyper spectral feature extraction through multiple hidden layers, showcasing a sophisticated and effective methodology for yield estimation.

- Robust Evaluation Metrics and Validation: Rigorous evaluation using metrics like R2, RMSE, MAE, and cosine similarity underscores the robustness and accuracy of the proposed Optimized CNN-LSTM model. The validation of the methodology's dependability provides confidence in the reliability of the findings, contributing to the advancement of agricultural remote sensing and predictive modelling.

Dataset In summary, the methodology combines advanced techniques from remote sensing, deep learning, and optimization to provide a comprehensive and effective approach for crop yield prediction and agricultural assessment. The contributions of this study have the potential to significantly impact precision agriculture by enabling farmers to make data-driven decisions for resource management and crop production. The remainder of this work is structured as follows: Section II offers a full analysis of these as well as related previous work. Details of the problem statement are included in Section III. In Section IV, the suggested Optimized CNN-LSTM architectures are covered in more depth. The results of the trials are discussed in Section V, and the proposed technique is thoroughly compared with existing best practices. Section VI concludes the paper.

II. Related Works

A predicted scientific approach is the integration of self-governing computing and artificial intelligence technology for agricultural ideas. With its extensive coverage, great spectral resolution, and wide range of narrow-band selection, the aerial hyperspectral system is a fantastic instrument for predicting crop physiological parameters and yield. It has been difficult to spread awareness of this technology due to the substantial and redundant three-dimensional analysis and computing. Based on three crop classifications with multi-functional cultivation, this research included two significant publicly available systems (R and Python), automatic hyperspectral narrow-band vegetation index estimation, and the most advanced machine learning (ML) modern equipment to calculate yield. Li et al.[14] demonstrated that AutoML regression model's predicted capacity was considerable. For single variety planting wheat, the best determination coefficient and normalized root mean square error (NRMSE) were 0.96 and 0.12, correspondingly; The restriction of the Auto-Sklearn approach, which prevents the investigation of the relevance ranking of individual feature, limits the capacity to retrace all regressors in this study, which may have an influence on the choice of appropriate vegetation indices.

Alfalfa is an important farmed feed crop. Due to effectiveness in data collecting, unmanned aerial vehicles (UAVs) are attracting in precision agriculture. hyperspectral data can provide a better level of spectral fidelity compared to other imaging techniques. Feng et al. [15] used UAV-based hyperspectral images, a feature selection is conducted to diminish the size of the data and retrieved a significant number of hyperspectral indices of the original image. Then, by merging three frequently used base learners, namely, support vector regression (SVR), K-nearest neighbors (KNN) and random forest (RF), an ensemble machine learning model was created. It demonstrated that ensemble model outperforms all base learners, and when employing the chosen features, an R_2 of 0.874 was obtained. The outcomes further support the effectiveness of the suggested ensemble model. The performance of the model might be affected by variables that were not taken into account for this research, such shifting field conditions, climatic variations, or insect infestations.

Artificial intelligence has easily migrated into a number of economic sectors, particularly for surveillance and control in agriculture. One of the main factors reducing crop output is biotic stress. Albanese, Nardello, and Brunelli [16] proposed Automatic recognition of hassle using images has emerged as a key study area for timely crop disease diagnosis by the advent of deep learning technologies. In order to continuously identify infestations of pests inside fruit orchards. The embedded approach is built on sensor device. The platform's capabilities have been shown through the training and deployment of three distinct ML algorithms. Furthermore, the incorporation of energy-harvesting functions into the suggested system guarantees extended battery lifetime. One drawback of this research is that the energy harvesting system's effectiveness heavily relies on the availability of sufficient sunlight, making it less practical in regions with limited sunlight or during extended periods of overcast weather.

Weed growth out of control has a negative impact on crop quality and productivity. Herbicide usage to eradicate weeds changes biodiversity and pollutes the environment. Instead, pinpointing weed-infested areas can help with targeted chemical remediation of these areas. There are now ways to detect weed plants due to improvements in agricultural image analysis. Supervised learning techniques needs a massive volume of human annotated images. Because there are so many different plant species being grown, these supervised systems are therefore economically unviable for the small-scale farmer. In this research, Shore Wala et al. [17] present a semi-supervised deep learning method. This weed quantity and distribution may be helpful in autonomous robot-assisted targeted treatment of diseased regions. Convolutional Neural Network (CNN) is used to identify the foreground vegetation indices including crops and weeds. The requirement for manually developing features is therefore removed by utilising a fine-tuned CNN to identify the weed-infected areas. The method is tested on two datasets (1) images of carrot plants, and (2) the Sugar Beets dataset. The suggested approach has an optimal recall of 0.99 and an average accuracy of 82.13% for estimating weed density in weed-infested areas. The limitation of this work is time consuming.

Techniques for proximal sensing may be used to examine soil and crop factors that affect agricultural output. By combining this precision agricultural technology with cutting-edge data processing techniques like machine learning (ML) algorithms, it is possible to fully realize their promise for managing crop productivity. In order to forecast potato tuber yield, four machine learning (ML) algorithms, namely elastic net (EN), linear regression (LR), and support vector regression (SVR), k-nearest neighbor (k-NN) were employed by Abbas et al. [18]. Data of soil were sampled for soil chemistry, moisture content of the soil and normalized-difference vegetative index (NDVI). At the conclusion of the growing season, manual data collection and yield sample collection took place. The data from three fields were then combined to create four datasets, PE-2017, NB-2018, NB-2017 and PE-2018which reflect the provincial data for the corresponding years. To develop yield projections evaluated using various statistical factors, modelling approaches were used. SVR models excelled all other models. The performance can be improved by using deep learning methods.

Crop supervision is changing as a result of neural networks and self-driving computers being integrated into agriculture. Crop production predictions are more accurately made thanks to overhead hyperspectral sensors and sophisticated predictive methods like combined models and AutoML regression. The accuracy of farming is improved by drones that operate using hyperspectral data; lucerne cultivation is one example of this. Using picture recognition and sensor devices, deep learning technologies facilitate the prompt identification of agricultural diseases. Promising results have been obtained using a semi-supervised deep learning approach to detect weeds, and machine learning methods are used to predict the yield of potato tubers depending on crop and soil characteristics. Even though these developments solve important issues, issues like consuming time and dependent on the climate energy collecting systems are still recognized.

III. Problem Statement

The existing methods may face limitations in feature ranking, energy availability for sensor devices [16], economic viability for small-scale farmers, time-consuming processes [17] and inefficient. To address these challenges, the proposed approach combines self-governing computing, hyperspectral feature extraction via Convolutional Neural Networks (CNN), and machine learning optimization techniques. Optimized CNN-LSTM model is introduced, integrating hyperspectral data analysis and yield prediction through multiple hidden layers. The model enhances performance through Firefly Algorithm optimization. This comprehensive approach contributes to accurate crop yield prediction, thereby improving precision agriculture practices.

IV. Methodology

This research capitalizes on Unmanned Aerial Vehicles (UAVs) equipped with hyperspectral sensors to meticulously capture intricate spectral data across agricultural terrains. The reliability and usability of the created models in actual agricultural contexts are ensured by modifying testing techniques to account for temperature, humidity, precipitation, and other meteorological parameters. Data preprocessing encompasses radiometric calibration and geometric corrections via the HyperSpec program, ensuring the precise conversion of original digital data into radiance data. Using NFINDR algorithm, mixed spectral signals in hyperspectral images are separated, revealing distinct fingerprints associated with crops, soil types, and agricultural variables. Integral to precision agriculture, a Convolutional Neural Network (CNN) framework extracts key features from hyperspectral data, including NDVI, CCCI, CVI, contrast, and entropy indices, heightening insights into crop health and landscape. A pioneering Optimized CNN-LSTM model is introduced for accurate crop yield prediction. Fusing deep learning employed with Firefly Algorithm optimization, the model integrates hyperspectral feature extraction, enabling precise yield estimation through multiple hidden layers. Rigorous evaluation, incorporating metrics like R^2, RMSE, MAE, and cosine similarity, underscores the Optimized CNN-LSTM model's robustness and accuracy, thus validating the proposed methodology's dependability. It is illustrated in Fig. 2.

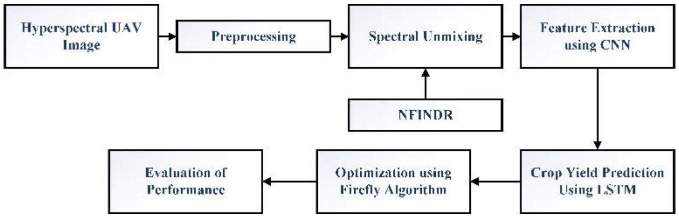

Hyperspectral UAV Image
Preprocessing
Spectral Unmixing
Feature Extraction using CNN
NFINDR
Crop Yield Prediction Using LSTM
Optimization using Firefly Algorithm
Evaluation of Performance

Fig. 2. Proposed model CNN-LSTM.

A. Data Collection and Preprocessing

Sensors attached on UAV flying over the countryside take hyperspectral images. Each pixel in the images has specific spectral characteristics [19]. The maker of the sensor used the HyperSpec program to perform radiometric calibrating and geometrical corrections on the UAV hyperspectral imagery. The sensor's laboratories evaluation factors, which may be represented as in Eq. (1), were used to convert the original digital number data into radiance data for the radiometric evaluation.

$$D_N = (R_1G_1 + R_2G_2) \times T_E + F_D \tag{1}$$

where, R_1 is the anticipated radiometric standard radiance for the first-order reflected image, R_2 is the radiometric standard radiance for the second-order reflected image, R_1 and R_2's system gains are G_1 and G_2, while the sensor's exposure time T_E is and the darkness of field measurement is F_D. Since the R2 strength is so poor, a spectral filter may be used to filter it. In most cases, the calibration variables are determined by the producer using an integrated sphere in the lab, and they are then stored in the calibration program. According to a correlated equation, GPS (Global Positioning System) and also inertia measuring units (IMU) modules' location and orientation data, and other data, the geometric correction can be carried out. The input parameters for the HyperSpec program include the original hyperspectral information, its frame indices, timestamps, digital surface model (DSM) information, and the GNSS/IMU files containing yaw, pitch, roll, latitude, and longitude [20]. Further enhancing the precision of the geometric correction may be done by using ground control points (GCPs).

B. Spectral Unmixing using NFINDR Algorithm

The process of "spectral unmixing" is employed to separate the mixed spectrum signals generated by hyperspectral images to its constituent parts. This entails recognizing and isolating the spectral fingerprints of numerous factors in relation to farming, such as distinct crops, different kinds of soil, and sometimes even pests or illnesses that may be detected in the fields. The landscape's structure will benefit from knowing that data. Five distinct groups, containing soil, shadows, spikes, crop leaves and gray panels, were present in hyperspectral imagine data. Thus, each of these five groups can all contribute to the spectral sensitivity of a pixel to varying degrees and with somewhat different spectral signatures. Five endmembers, every indicating one group, were found from the images taken at five-meter height to show the large number of these categories of each pixel found in the images obtained at 20-meter height. When these images are made, the ensuing dataset mimics one that can be logically collected anyway. The NFINDR technique is used to execute spectral unmixing at the subsequent step. The NFINDR approach is essentially an automated method for locating the cleanest pixels in a picture. The primary goal of this approach is to precisely replicate the effective (non-automated) method of locating the highest and lowest points of scatter chart. The convex structure of the available hyperspectral data makes it possible to use the NFINDR methodology in a reasonably simple and rapid manner. Endmembers = nfindr (inputData, numEndmembers, Name, Value) is a built-in MATLAB code that has been used in the suggested workflow. This function uses the NFINDR technique to obtain endmember signals out of hyperspectral data. The amount of endmember signatures to be retrieved with the NFINDR technique is denoted by numEndmembers. This form is employed when both reduction of dimensionality and parameters for the number of cycles are necessary. The extracted endmembers relate to certain soil constituents at specified wavelengths. For a variety of uses, such as crop detection and tracking in agriculture utilizing UAV-based hyperspectral imaging, the NFINDR technique has shown to be excellent for unmixing hyperspectral data efficiently on the computer [21].

C. NFINDR Algorithm

1) Minimize the incoming data's spectral dimensionality and estimate the primary component bands. Decide the quantity of PC wavelengths and endmembers will be extracted, and then extract that amount.

2) As a first group of endmembers, randomly select n pixel spectrum out of the minimized data.

3) Initialize iteration 1 and first collection of endmembers to determine the volume using $V\left(M^{(1)}\right) = |\det(M^{(1)})|$

where, $M^{(1)} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ m_1^{(1)} & m_2^{(1)} & \ldots & m_p^{(1)} \end{bmatrix}$

4) Choose a new pixel spectrum s, for this second cycle., such that: $s \notin \left\{m_1^{(1)}, m_2^{(1)}, \cdots m_p^{(1)}\right\}$.

5) Calculate the area of the resultant simplex $V(M^{(2)})$ after replacing all endmember of the collection with r.

6) If the calculated volume $V(M^{(2)})$ is higher than $V(M^{(1)})$, update the ith endmember of the collection with r. A revised list of endmembers generated.

7) Choose a different pixel spectrum for each subsequent iteration, then repeat steps 5 and 6 with that new spectrum. Once the overall level of iterations reaches the desired number, the iterations come to an end.

D. Accuracy Prediction using Cosine Similarity

The accuracy for the spectral unmixing is assessed using cosine similarity as in Eq. (2), as well as it is also determined if the endmember frequencies acquired can be coincided with to the wavelengths originally utilized to create the hyperspectral images.

$$\text{Similarity=} \cos(\theta) = ((\mathrm{E} * \mathrm{P}) / (\|\mathrm{E}\| * \|\mathrm{P}\|)) \tag{2}$$

where, || || stands for the vector's magnitude (Euclidean norm), P is the vector representing each pixel's spectrum signature, and E is the chosen endmember option vector.

E. Extracted Features of Hyper Spectral Images using Convolutional Neural Network (CNN)

Precision agriculture, which applies resources like water, fertilizer, and pesticides especially when and where they are required, depends on the ability to extract features of hyperspectral crop images. Farmers may use resources more efficiently, cut down on waste, and boost crop output by having a better grasp of the variation in crop nutritional demands throughout a field. The features were extracted using CNN.

1) Normalized Difference Vegetation Index (NDVI): NDVI represents the health and vigor of plants in an area. It is estimated as of the reflectance of two bands, typically the near-infrared (IR_N) and the red (R) bands, using the following Eq. (3).

$$NDVI = \frac{IR_N - R}{IR_N + R} \tag{3}$$

NDVI values range from -1 to +1, where positive values point to healthy vegetation, zero represents non-vegetated areas (e.g., water bodies), and negative values (closer to -1) typically represent cloud cover or man-made surfaces.

2) Canopy Chlorophyll Content Index (CCCI): CCCI is another vegetation index used to estimate the chlorophyll content in vegetation canopies. It is particularly useful for monitoring the greenness and health of vegetation. The CCCI involves the red-edge band (R_E) is given in Eq. (4).

$$CCCI = \frac{\frac{IR_N - R_E}{IR_N + R_E}}{NDVI} \tag{4}$$

Higher CCCI values indicate higher chlorophyll content and healthier vegetation.

a) Chlorophyll VI (CVI): Chlorophyll VI is an index designed to provide a more accurate estimation of chlorophyll content in vegetation as in Eq. (5). It utilizes the red band (R), and the green band (G).

$$CVI = IR_N \times \frac{R}{G^2} \tag{5}$$

CVI values are positively correlated with higher chlorophyll content.

b) Contrast: Contrast is a statistical measure used to describe the difference in pixel intensities within an image or a specific region. In hyperspectral imagery, contrast refers to the variation in spectral values across different bands. Higher contrast indicates more pronounced differences between spectral signatures, which can be useful for discriminating between different materials or classes as shown in Eq. (6).

$$contrast = \sum_{k,l=1}^{N} P_{k,l}(k-l)^2 \tag{6}$$

c) Entropy: Entropy is another statistical measure that quantifies the amount of uncertainty or disorder in an image or a specific region as in Eq. (7). In hyperspectral images, entropy can be used to assess the complexity and variability of spectral signatures. High entropy values indicate greater spectral diversity and complexity, which can be helpful for identifying diverse land cover types.

$$Entropy = \sum_{k,l=1}^{N} P_{k,l} (\ln P_{k,l})^2 \tag{7}$$

NDVI, CCCI, and CVI are all vegetation indices that provide information about the health and vigor of crops. By calculating these indices from hyperspectral data, it becomes possible to monitor the growth status, stress levels, and overall health of crops [22].

F. Yield Estimation using LSTM

LSTM [23] By including a gradient superhighway in the structure of a state of a cell c as well to the hidden state h, a unique type of RNN was developed to address this problem. The LSTM architecture features gates that allow both the addition and removal of data from the cell state. The forget gate determines whether data should be eliminated from the current state of the cell and is described as follows:

$$f'_{t=\sigma\left(W'^{T}_{f'}\left[h'_{t-1,x'_t}\right]+b'_f\right)} \tag{8}$$

The definition of the input gate that chooses the data to be fed to the state of the cell is in Eq. (9).

$$i'_{t=\sigma(W'^{T}_{i^{i'}}\left[h'_{t-1,x'_t}\right]+b'_{i'})} \tag{9}$$

Utilizing both and it, the cell state c'_t is derived in the way shown in Eq. (10).

$$\dot{c}'_{t=tanh\left(W'^{T}_{c'}\left[h'_{t-1,x'_t}\right]+b'_c\right)} \tag{10}$$

$$c'_{t=\tanh(}f'^{T}_{t}\,c'_{t-1+}i'^{T}_{t}\,\dot{c}_t \tag{11}$$

The concealed and outgoing state o'_t of the LSTM, respectively, are specified as.

$$o'_{t=\sigma(W'^{T}_{0}\left[h'_{t-1,x'_t}\right]+b'_0)} \tag{12}$$

$$h_t = o'^{T'}_{t}\tanh(c_t) \tag{13}$$

Fig. 3 depicts the architecture of proposed CNN-LSTM model. Due to a more efficient gradient flow during back propagation, LSTM is more successful at simulating lengthy sequences than a straightforward RNN.

G. Optimization using Firefly Algorithm

An optimization approach was utilized to determine the parameters for which the cost function was minimized throughout the training procedure. The variation in light intensity and the evolution of attraction serves as the two pillars of the firefly optimization approach. A measure of attraction is the intensity, which is connected to the objective function. The relative attractiveness (α) as judged by other fireflies fluctuates when the distance d_{ij} between fireflies i and j shifts. Light loses intensity as it moves farther away from its source due to air absorption. Additionally, as shown in Eq. (14), the attractiveness varies according to the degree of absorption and the brightness I(d) varies in accordance with the inverse square law.

$$I(d) = \frac{I_s}{d^2} \tag{14}$$

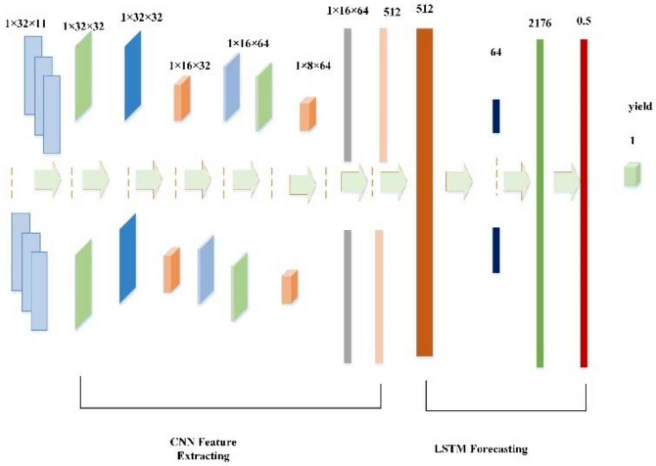


Fig. 3. The architecture of proposed CNN-LSTM model.

where, d is the distance between the source and the light, δ is the absorption coefficient of light, and I_s is the source intensity. Eq. (15) states that the light's intensity (I), which is dependent on its coefficient of absorption, varies with distance d.

$$I = I_0 e^{-\delta d^2} \tag{15}$$

where,,I_0 is the initial brightness of the light, and $e^{-\delta d^2}$ is the sum of the intensities at the source and at a distance. The chosen attributes that each firefly location represents are used to assess the effectiveness of the crop Yield detecting model. Each firefly's attractiveness (α) in comparison to others is determined based on the fitness values received during the evaluation, as stated in Eq. (16).

$$\alpha = \alpha_0 e^{-\delta d^2} \tag{16}$$

Higher fitness values make fireflies attractive. After that, transport the firefly in that direction so they can explore the search area. To strike a balance between exploration and exploitation, adjust the Firefly Optimization algorithm's control parameters, such as the attraction coefficient, absorption coefficient, and iterations. Decide on the termination criteria, which govern when the optimization process should be stopped. Reaching a predetermined number of iterations or obtaining a good fitness value are frequent reasons for termination [24].

Algorithm 1: Firefly Algorithm

```
Initialize fireflies randomly
Set control parameters (alpha, delta, iterations)
Repeat for a predefined number of iterations:
 Calculate fitness values for each firefly
 Update the attractiveness (alpha) based on fitness and distance
 Move fireflies towards more attractive ones
 Explore the search space
Until termination criteria met
```

V. Result and Discussion

Following is an explanation of how to test the accuracy of the CNN-LSTM model using the coefficient of determination (R2), mean absolute deviation (MAE), and root mean square error (RMSE):

A. *Model Accuracy Assessment Parameters*

To assess the accuracy of the yield prediction approach, the coefficient of determination (R^2), mean absolute deviation (MAE), and root mean square error (RMSE) were utilized. R^2, MAE, and RMSE are calculated as in Eq. (17), Eq. (18) and Eq. (19).

$$R^2 = 1 - \frac{\sum_{k=1}^{N}(\hat{y}_k - y_k)^2}{\sum_{k=1}^{N}(y_k - \bar{y})^2} \tag{17}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(\hat{y}_k - y_k)^2}{N}} \tag{18}$$

$$MAE = \frac{1}{k}\sum_{k=1}^{N}|y_k - \widehat{y_k}| \tag{19}$$

where, $\bar{y}$ is the mean for the detected crop yield, y_k and $\hat{y}_k$ are the observed and estimated crop yields, respectively; N represents the number of evaluation samples. A greater prediction performance of the model is shown by an increased R^2 and a decreased RMSE [25].

The assessment parameters for the Optimized CNN-LSTM model in Table I demonstrate its exceptional performance. The hyperspectral image dataset's underlying patterns are very well captured and understood by the model, as seen by the Coefficient of Determination (R2) value of 0.893, demonstrating the model's capacity to explain variations in observed data. Furthermore, the model's accuracy and dependability in predicting crop yields are shown by the low Root Mean Square Error (RMSE) of 0.13 and Mean Absolute Error (MAE) of 0.14. These findings demonstrate Optimized CNN-LSTM potential as a formidable tool for hyperspectral image-based yield forecasting, providing insightful information for remote sensing and precision agricultural applications. Fig. 4 depicts the Assessment parameter of Optimized CNN-LSTM.

B. *Statistical Analysis*

To investigate the impact of variation in breeds, irrigation methods, and their relations on the observed and anticipated crop yield, a mixture of linear model was used. The model's equation is given in Eq. (16).

$$Y = \alpha X + \gamma Z + e \tag{16}$$

where, X and Z stand for static effects and random effects accordingly, Y is the response shown by fixed effect (α) as well as random effect (γ) by a random error (e). With an interval from zero to one, broad-sense heritability measures the proportion of genetic variance to all phenotypic variation. The variance in phenotype is totally influenced by genetic and environmental variables, respectively, as indicated by the heritability of 0 and 1. The following Eq. (17) was used to compute the heritability.

$$h^2 = {v_g}\Big/{\left(v_g + {v_e}/{r}\right)} \tag{17}$$

where, v_g and v_e stand for the genetic and erroneous variances, correspondingly, and r stands for the number of reproductions per treatment.

Table II presents the Coefficient of determination (R^2) performance metrics for different predictive models. Random Forest (RF) achieved an R^2 value of 0.882, indicating a strong ability to explain the variance in the data. Support Vector Regression (SVR) performed well with an R^2 of 0.824, capturing a substantial portion of the data's variability. Optimized CNN-LSTM model exhibited the highest performance, attaining an R^2 of 0.893, signifying its superior capability in predicting and understanding the underlying patterns within the dataset. Fig. 5 shows comparison of various models with R^2.

TABLE I. Assessment Parameter of Optimized CNN-LSTM

Assessment Parameters	Values
R2	0.893
RMSE	0.15
MAE	0.14

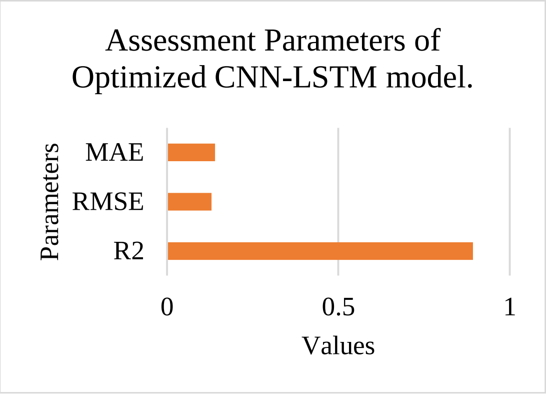


Fig. 4. Assessment parameter of optimized CNN-LSTM.

TABLE II. Accuracy Assessment Parameter

Model	R^2
RF	0.882
SVR	0.824
Optimized CNN-LSTM	**0.893**

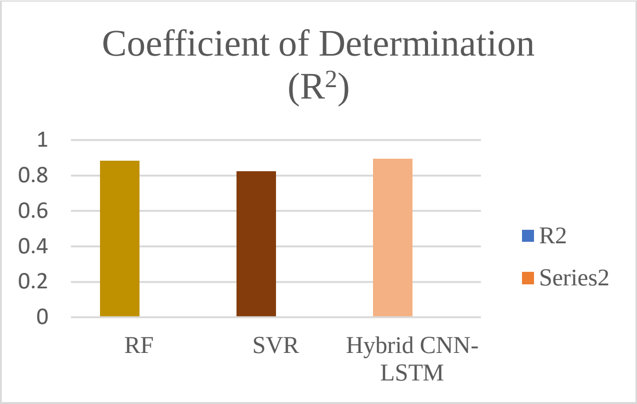


Fig. 5. Comparison of R^2 for various models.

Fig. 6 which depicts the training accuracy of proposed Optimized CNN-LSTM model. The accuracy model is plotted based on the value of R^2. It clearly shows that the proposed method outperforms other algorithms.

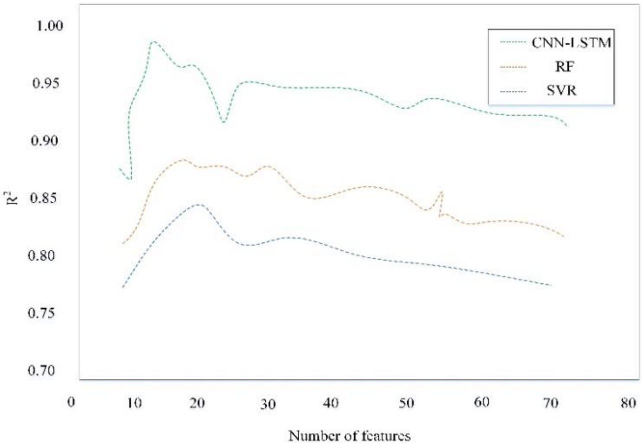


Fig. 6. Training accuracy with existing model.

This research used a dataset obtained from Unmanned Aerial Vehicles (UAVs) to analyze hyper spectral images in agricultural areas. The dataset was divided into two subsets for model development and evaluation. 80% of the dataset was allocated for the training phase, where the model learns patterns and relationships from the hyper spectral data. The remaining 20% was used for the testing phase, where the model encounters new data and evaluates its performance. This division is a standard practice in machine learning to gauge performance, prevent over fitting, and ensure reliability in real-world applications. The careful consideration of dataset partitioning is crucial for establishing the effectiveness and generalization of the developed models in agricultural hyper spectral image analysis. The training and test dataset were illustrated in Fig. 7.

The NFINDR algorithm is utilized to perform spectral unmixing on hyperspectral images. This technique separates mixed spectral signals into their constituent parts, identifying and isolating spectral fingerprints associated with different crops, soil types, and other factors relevant to agriculture. End members were selected as shown in Fig. 8.

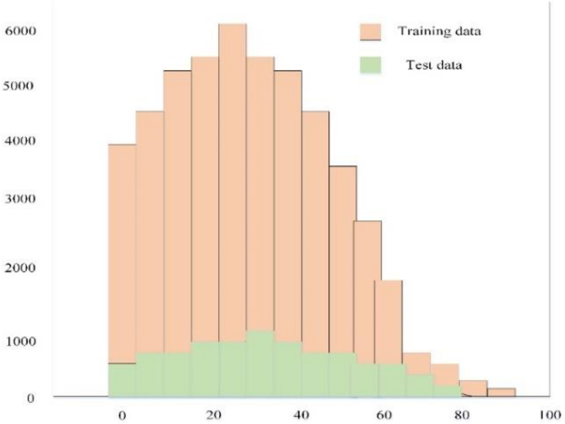


Fig. 7. Training and test dataset.

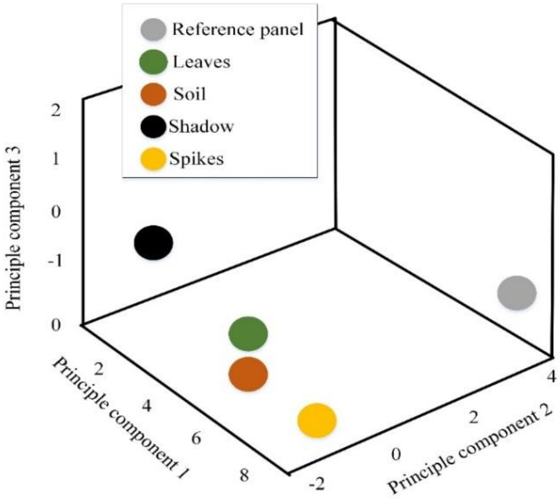


Fig. 8. Spectral unmixing using end members selection in NFINDR.

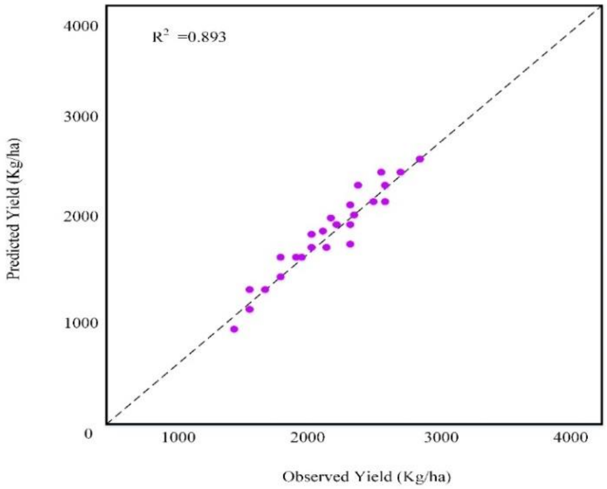


Fig. 9. The relation between observed and predicted yield with R^2.

Fig. 9 depicts the relationship between observed and predicted yields with a coefficient of determination (R^2) value of 0.89. The data points closely follow a linear trend, indicating a link between the predicted and actual yields. The high R^2 value suggests that approximately 89% of the variability in the observed yield can be explained by the predictive model, affirming its accuracy and reliability in forecasting agricultural yields. This alignment between predicted and observed values underscores the model's effectiveness in capturing the underlying patterns and factors influencing crop production.

Table III presents Root Mean Square Error (RMSE) values for different predictive models. The Random Forest (RF) model yielded an RMSE of 0.22, reflecting the average magnitude of prediction errors in relation to the observed values. The Support Vector Regression (SVR) model produced a slightly higher RMSE of 0.23, indicating a slightly greater overall error compared to RF. In contrast, the Optimized CNN-LSTM model demonstrated the lowest RMSE at 0.13, signifying its superior accuracy in predicting outcomes and minimizing prediction errors. These RMSE values provide insights into the precision and effectiveness of each model, with Optimized CNN-LSTM model. model emerging as the most precise in this analysis. The Training and testing loss is plotted based on RMSE value.

One of the most important aspects of this study's model evaluation is the training and validation loss analysis. The model's learning process and its capacity to generalize to new data are revealed by the plot of training and validation loss based on the Root Mean Square Error (RMSE) value. The model is effectively learning from the training data, indicating a successful training procedure, as seen by the decreasing trend in both training and validation loss. Given that there is little substantial divergence between the training and validation loss curves, it is possible that the model is not over fitting. This alignment shows that the model may generalize effectively to new, unexplored data points, increasing its dependability in real-world situations.

The model's accuracy and precision in forecasting crop yields are further supported by the low RMSE values linked to the training and validation loss, securing its application in remote sensing and precision agriculture settings. The training and testing loss expressed as RMSE is shown in Fig. 10.

Table IV showcases Mean Absolute Error (MAE) values for different predictive models. The Random Forest (RF) model achieved an MAE of 0.17. The Support Vector Regression (SVR) model exhibited a slightly lower MAE of 0.16, indicating a marginally improved accuracy in prediction compared to RF. Remarkably, Optimized CNN-LSTM model. Model outperformed the others with the lowest MAE of 0.14, underscoring its exceptional precision in forecasting and its ability to minimize absolute prediction errors. These MAE values provide valuable insights into the predictive capabilities of each model, with the Optimized CNN-LSTM model. It demonstrates the highest level of accuracy in this context.

The crop yield prediction models' accuracy was assessed using the accuracy assessment metrics coefficient of determination (R2), mean absolute deviation (MAE), and root mean square error (RMSE). The effectiveness of several prediction models, such as Random Forest (RF), Support Vector Regression (SVR), and Optimized CNN-LSTM model. It was evaluated using these measures. The outcomes showed that the Optimized CNN-LSTM model. It performed better than the competing models, earning the greatest R2 value of 0.893, demonstrating its improved capacity to account for the variance in the crop production data. Optimized CNN-LSTM model further demonstrated the lowest RMSE and MAE values at 0.13 and 0.14, respectively, emphasizing its outstanding accuracy in predicting crop yields. These findings highlight the value of combining deep learning, spectrum unmixing, and hyperspectral imaging in precision agriculture, with Optimized CNN-LSTM model demonstrating the most promising outcomes. The observed and projected yield connection, which has an R2 value of 0.89 and indicates that almost 89% of the variability in observed yield can be explained by the predictive model, provided more evidence of the model's accuracy. With useful insights for large-scale farming operations, this research shows the potential of cutting-edge technology and machine learning approaches to improve crop output estimates.

TABLE III. ROOT MEAN SQUARE ERROR (RMSE)

Model	RMSE
RF	0.22
SVR	0.23
Optimized CNN-LSTM	**0.15**

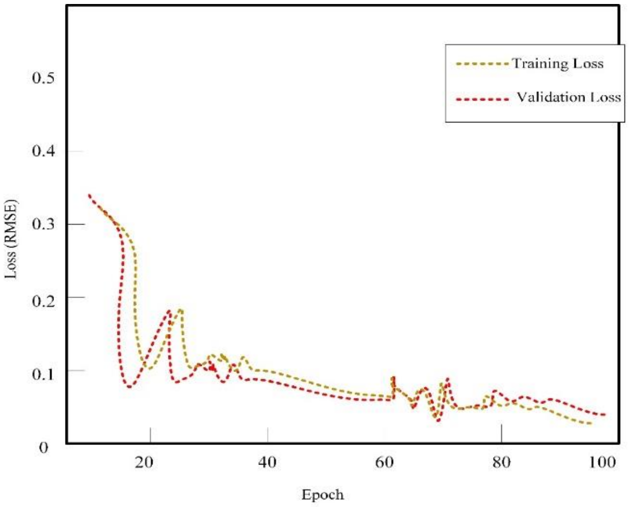


Fig. 10. Training and validation loss.

TABLE IV. MAE VALUES COMPARISON

Model	MAE
RF	0.17
SVR	0.16
Optimized CNN-LSTM	**0.14**

C. Discussion

The integration of self-governing computing and artificial intelligence in agriculture, particularly in crop monitoring and yield prediction, shows promise. Utilizing aerial hyper spectral systems and UAVs, combined with advanced machine learning techniques, improves crop yield predictions. Examples include AutoML regression models for wheat and an ensemble model for alfalfa. Deep learning technologies are used for disease identification and semi-supervised weed detection. Despite challenges like time-consuming processes and climate-dependent energy-harvesting systems, the adoption of neural networks and precision agriculture technologies is transforming crop management.

VI. Conclusion and Future Work

The suggested model combines deep learning, spectral unmixing, and hyper spectral imaging in a way that improves upon earlier methods for projecting crop production. As opposed to traditional methods that depend on subjective evaluations, this new approach makes use of hyper spectral photography to gather copious amounts of non-visible data, which enables a thorough analysis of crop characteristics. Spectral unmixing methods allow for the accurate assessment of crop health and growth stages by identifying distinct spectral signatures. This improved spectral data is then used by deep learning algorithms to create a solid, data-driven basis for precise crop production forecasts. A model with the lowest Root Mean Square Error (RMSE) of 0.15 and Mean Absolute Error (MAE) of 0.14 is produced by integrating these cutting-edge approaches with MATLAB, demonstrating improved prediction accuracy over existing models. This novel strategy overcomes the drawbacks of traditional approaches and greatly improves crop output forecast capabilities, marking a significant advance in precision agriculture. Future research in precision agriculture ought to concentrate on improving data integration by investigating the fusion of various sensors, IoT devices, and remote sensing technologies to capture a thorough dataset for crop monitoring. In order to provide openness and confidence in yield projections, explainable AI models also need to be developed.

References

[1] M.-L. Tseng, A. S. F. Chiu, C.-F. Chien, and R. R. Tan, "Pathways and barriers to circularity in food systems," Resour. Conserv. Recycl., vol. 143, pp. 236–237, Apr. 2019, doi: 10.1016/j.resconrec.2019.01.015.

[2] P. K. Sethy, C. Pandey, Y. K. Sahu, and S. K. Behera, "Hyperspectral imagery applications for precision agriculture - a systemic survey," Multimed. Tools Appl., vol. 81, no. 2, pp. 3005–3038, Jan. 2022, doi: 10.1007/s11042-021-11729-8.

[3] G. Avola, A. Matese, and E. Riggi, "An Overview of the Special Issue on 'Precision Agriculture Using Hyperspectral Images,'" Remote Sens., vol. 15, no. 7, p. 1917, Apr. 2023, doi: 10.3390/rs15071917.

[4] T. B. Shahi, C.-Y. Xu, A. Neupane, and W. Guo, "Machine learning methods for precision agriculture with UAV imagery: a review," Electron. Res. Arch., vol. 30, no. 12, pp. 4277–4317, 2022, doi: 10.3934/era.2022218.

[5] N. Sulaiman, N. N. Che'Ya, M. H. Mohd Roslim, A. S. Juraimi, N. Mohd Noor, and W. F. Fazlil Ilahi, "The Application of Hyperspectral Remote Sensing Imagery (HRSI) for Weed Detection Analysis in Rice Fields: A Review," Appl. Sci., vol. 12, no. 5, p. 2570, Mar. 2022, doi: 10.3390/app12052570.

[6] M. Ruiz, M. J. Beriain, M. Beruete, K. Insausti, J. M. Lorenzo, and M. V. Sarriés, "Application of MIR Spectroscopy to the Evaluation of Chemical Composition and Quality Parameters of Foal Meat: A Preliminary Study," Foods, vol. 9, no. 5, p. 583, May 2020, doi: 10.3390/foods9050583.

[7] H. J. Escalante, S. Rodríguez-Sánchez, M. Jiménez-Lizárraga, A. Morales-Reyes, J. De La Calleja, and R. Vazquez, "Barley yield and fertilization analysis from UAV imagery: a deep learning approach," Int. J. Remote Sens., vol. 40, no. 7, pp. 2493–2516, Apr. 2019, doi: 10.1080/01431161.2019.1577571.

[8] B. Verma, R. Prasad, P. K. Srivastava, P. Singh, A. Badola, and J. Sharma, "Evaluation of Simulated AVIRIS-NG Imagery Using a Spectral Reconstruction Method for the Retrieval of Leaf Chlorophyll Content," Remote Sens., vol. 14, no. 15, p. 3560, Jul. 2022, doi: 10.3390/rs14153560.

[9] B. Lu, P. Dao, J. Liu, Y. He, and J. Shang, "Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture," Remote Sens., vol. 12, no. 16, p. 2659, Aug. 2020, doi: 10.3390/rs12162659.

[10] C. Sun et al., "Prediction of End-Of-Season Tuber Yield and Tuber Set in Potatoes Using In-Season UAV-Based Hyperspectral Imagery and Machine Learning," Sensors, vol. 20, no. 18, p. 5293, Sep. 2020, doi: 10.3390/s20185293.

[11] J. Nalepa, "Recent Advances in Multi- and Hyperspectral Image Analysis," Sensors, vol. 21, no. 18, p. 6002, Sep. 2021, doi: 10.3390/s21186002.

[12] S. A. Bhat and N.-F. Huang, "Big Data and AI Revolution in Precision Agriculture: Survey and Challenges," IEEE Access, vol. 9, pp. 110209–110222, 2021, doi: 10.1109/ACCESS.2021.3102227.

[13] C. Nguyen, V. Sagan, M. Maimaitiyiming, M. Maimaitijiang, S. Bhadra, and M. T. Kwasniewski, "Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning," Sensors, vol. 21, no. 3, p. 742, Jan. 2021, doi: 10.3390/s21030742.

[14] K.-Y. Li et al., "Toward Automated Machine Learning-Based Hyperspectral Image Analysis in Crop Yield and Biomass Estimation," Remote Sens., vol. 14, no. 5, p. 1114, Feb. 2022, doi: 10.3390/rs14051114.

[15] L. Feng et al., "Alfalfa Yield Prediction Using UAV-Based Hyperspectral Imagery and Ensemble Learning," Remote Sens., vol. 12, no. 12, p. 2028, Jun. 2020, doi: 10.3390/rs12122028.

[16] A. Albanese, M. Nardello, and D. Brunelli, "Automated Pest Detection with DNN on the Edge for Precision Agriculture," IEEE J. Emerg. Sel. Top. Circuits Syst., vol. 11, no. 3, pp. 458–467, Sep. 2021, doi: 10.1109/JETCAS.2021.3101740.

[17] S. Shorewala, A. Ashfaque, R. Sidharth, and U. Verma, "Weed Density and Distribution Estimation for Precision Agriculture Using Semi-Supervised Learning," IEEE Access, vol. 9, pp. 27971–27986, 2021, doi: 10.1109/ACCESS.2021.3057912.

[18] F. Abbas, H. Afzaal, A. A. Farooque, and S. Tang, "Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms," Agronomy, vol. 10, no. 7, p. 1046, Jul. 2020, doi: 10.3390/agronomy10071046.

[19] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," Remote Sens. Environ., vol. 250, p. 112012, Dec. 2020, doi: 10.1016/j.rse.2020.112012.

[20] A. Vinci, R. Brigante, C. Traini, and D. Farinelli, "Geometrical Characterization of Hazelnut Trees in an Intensive Orchard by an Unmanned Aerial Vehicle (UAV) for Precision Agriculture Applications," Remote Sens., vol. 15, no. 2, p. 541, Jan. 2023, doi: 10.3390/rs15020541.

[21] K. Lavanya, R. Jaya Subalakshmi, T. Tamizharasi, L. Jane, and A. Victor, "Unsupervised Unmixing and Segmentation of Hyper Spectral Images Accounting for Soil Fertility," Scalable Comput. Pract. Exp., vol. 23, no. 4, pp. 291–301, Dec. 2022, doi: 10.12694/scpe.v23i4.2031.

[22] W. Yang et al., "Estimation of corn yield based on hyperspectral imagery and convolutional neural network," Comput. Electron. Agric., vol. 184, p. 106092, May 2021, doi: 10.1016/j.compag.2021.106092.

[23] S. Sharma, S. Rai, and N. C. Krishnan, "Wheat crop yield prediction using deep LSTM model," ArXiv Prepr. ArXiv201101498, 2020.

[24] N. Kardani, A. Bardhan, P. Samui, M. Nazem, A. Zhou, and D. J. Armaghani, "A novel technique based on the improved firefly algorithm coupled with extreme learning machine (ELM-IFF) for predicting the thermal conductivity of soil," Eng. Comput., vol. 38, no. 4, pp. 3321–3340, Aug. 2022, doi: 10.1007/s00366-021-01329-3.

[25] S. Fei, L. Li, Z. Han, Z. Chen, and Y. Xiao, "Combining novel feature selection strategy and hyperspectral vegetation indices to predict crop yield," Plant Methods, vol. 18, no. 1, p. 119, Nov. 2022, doi: 10.1186/s13007-022-00949-0.

Optimizing Network Security and Performance Through the Integration of Hybrid GAN-RNN Models in SDN-based Access Control and Traffic Engineering

Ganesh Khekare[1], Dr.K.Pavan Kumar[2], Kundeti Naga Prasanthi[3], Dr. Sanjiv Rao Godla[4],
Venubabu Rachapudi[5], Dr. Mohammed Saleh Al Ansari[6], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[7]

Associate Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India[1]
Sr Asst.Prof, Dept.of IT, Prasad V Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada -07[2]
Dept.of CSE, Lakireddy Balireddy College of Engineering, Mylavaram[3]
Professor, Department of CSE (Artificial Intelligence & Machine Learning), Aditya College of Engineering & Technology,
Surampalem, Andhra Pradesh, India[4]
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram,
Guntur, Andhra Pradesh, India-522302[5]
Associate Professor, College of Engineering-Department of Chemical Engineering, University of Bahrain, Bahrain[6]
Faculty of Informatics and Computing, UniSZA University, Malaysia[7]

Abstract—**By offering flexible and adaptable infrastructures Software-Defined Networking (SDN) has emerged as a disruptive technology that has completely changed network provisioning and administration. By seamlessly integrating Hybrid Generative Adversarial Network-Recurrent Neural Network (GAN-RNN) modeling into the foundation of SDN-based traffic engineering and accessibility control methods, this work presents a novel and comprehensive method to improve network efficiency and security. The proposed Hybrid GAN-RNN models address two important aspects of network management: traffic optimization and access control. They combine the benefits of Generative Adversarial Networks (GANs) and Recurrent Neural Networks (RNNs). Traditional traffic engineering techniques frequently find it difficult to quickly adjust to situations that are changing quickly within today's dynamic networking environments. The models' capacity to generate synthetic traffic patterns that nearly perfectly replicate the complexity of real network traffic demonstrates the power of GANs. Network administrators can now allocate resources and routing methods more dynamically, as well as in responding to real-time network inconsistencies, due to this state-of-the-art technology. The technique known as Hybrid GAN-RNN addresses the enduring problem of network security. With their reputation for continuous learning and by utilizing Python software, recurrent neural networks (RNNs) are at the forefront of developing flexible management of access rules. With an incredible 99.4% accuracy rate, the "Proposed GAN-RNN" approach outperforms the other approaches. A comprehensive evaluation of network traffic and new safety risks allow for the immediate modification of these policies. This work is interesting because it combines hybrid GAN-RNN algorithms to strengthen security protocols with adaptive access control while also optimizing network efficiency through realistic traffic modeling.**

Keywords—Software-defined networking; generative adversarial networks; recurrent neural networks; traffic engineering

I. Introduction

Network performance directly impacts the efficiency of operations within an organization. Faster data transfer and lower latency lead to increased productivity, reduced downtime, and better user experiences, all of which are critical in today's fast-paced digital world [1]. Slow or unreliable networks result in poor user experiences. This can frustrate customers, employees, and partners, leading to dissatisfaction and potentially driving them away. Ensuring a high-performing network enhances user satisfaction and loyalty. In an era where data is a valuable asset, efficient network performance is crucial for transferring large volumes of data quickly and securely [2]. This is especially important for industries like healthcare, finance, and media, where sensitive information needs to be transmitted reliably. Cyber security threats are on the rise, and networks are prime targets for attacks. A secure network helps protect sensitive data, prevents unauthorized access, and mitigates risks associated with data breaches, which can have severe legal, financial, and reputational consequences. Many industries and organizations must adhere to strict regulatory compliance standards regarding data security and privacy. Maintaining a secure network is essential to meeting these requirements and avoiding legal penalties. A well-optimized network ensures that network resources, such as bandwidth and hardware, are used efficiently. This reduces costs associated with network maintenance and upgrades while maximizing resource availability [3]. Network failures or security breaches can disrupt business operations, leading to downtime and financial losses. Ensuring network resilience and security is crucial for business continuity and disaster recovery planning. As businesses grow, their network needs often grow too. A well-optimized network can scale to accommodate increased traffic, new devices, and expanding operations without sacrificing performance or security. Organizations with high-performing, secure networks can gain a competitive advantage [4]. They can offer faster services, better customer experiences, and innovative solutions that competitors with subpar networks may struggle to match. Improved network performance and security enable the adoption of advanced technologies like IoT (Internet of Things), cloud computing, and AI, which can drive innovation and digital transformation within an organization. Innovative solutions are required due to the constantly changing network infrastructure landscape in order to improve security and performance. SDN, or software-defined networking, has become a game-changing network management technology that gives network administrators the freedom to customize access control and traffic engineering [5]. In order to enhance network performance and security in SDN settings, this study investigates a unique method that combines hybrid generative adversarial networks (GANs) with recurrent neural networks (RNNs). The project intends to address important issues in traffic engineering and access control by smoothly integrating these cutting-edge machine learning approaches, leading to ultimately more effective and secure network operations [6]. The article digs into the principles of Software-Defined Networking (SDN) and discusses how it applies to contemporary network architecture. It gives an overview of how SDN can provide centralized network administration and dynamic traffic engineering by separating the control and data planes. As a prelude to the suggested remedy, the section also illustrates the difficulties in optimizing traffic flows in SDN settings [7].

The use of techniques based on machine learning has a lot of potential to enhance network performance in several ways. ML algorithms have the capacity to analyses network data, adjust to changing circumstances, and make choices in real time, which can improve network operations' efficiency, dependability, and security. ML models can spot unusual network activity that may indicate security risks or performance problems. Network assaults may be swiftly detected and responded to by intrusion detection systems (IDS) and intrusion prevention systems (IPS) driven by ML, improving security while minimizing service interruption [8]. Applications will perform better and use resources more effectively as a result of getting the resources they require when they need them. To maintain equitable server utilization, ML models may track server load and distribute requests that arrive around servers. Based on past data and user behavior, ML may help forecast future network requirements [9]. This can assist network administrators in making plans for infrastructure or capacity modifications so that the network is ready to meet growing demand. For instance, during periods of high demand, they might give priority to particular sorts of traffic. GANs use a loss function that guides the training process. The generator's loss depends on how well the discriminator is fooled, while the discriminator's loss is based on its ability to distinguish real from fake data. The training aim to find a balance where the generator generates highly convincing data and the discriminator becomes uncertain about its classifications. Access control in the context of SDN is covered in the second part. It explains the idea of dynamic access control lists and discusses the significance of access control for network security [10]. It examines the difficulties with access control in SDN, highlighting the requirement for more sophisticated and flexible security mechanisms.

The merging of Hybrid GANs and RNNs, the research's main novelty, is presented in the publication. It describes how RNNs may examine this data to find patterns and abnormalities using synthetic network traffic data produced by GANs. This hybrid strategy tries to simultaneously improve network security and performance. The practical use of the hybrid GAN-RNN models for traffic engineering is covered in this section. It offers information on how dynamic network traffic flow optimization may be accomplished using synthetic traffic data produced by GANs. The advantages of this strategy are explored in terms of decreased congestion, enhanced Quality of Service (QoS), and effective resource utilization. The research examines the use of hybrid GAN-RNN models in SDN access control. It explains how RNNs may inspect network traffic data for irregularities and security risks. The flexibility of this strategy to changing security threats is discussed, as is how it enhances access control by dynamically updating access lists in response to threat detection in real time. The article covers the overall effects of incorporating Hybrid GAN-RNN models into SDN settings, highlighting the enhancements to network security and performance. It also describes possible future avenues for study and application in the area of SDN-based traffic engineering and access management. The current limitation of these investigations is the lack of focus on scalability and real-world implementation, which makes it difficult to actually apply suggested security solutions in intricate network systems. Furthermore, there isn't much talk about possible interoperability issues, resource limitations, and how well these solutions may change to meet new threats in the cyberspace. A more thorough examination of these factors might improve the research findings' relevance and efficacy in real-world contexts.

Key contributions of the research include:

- Using traffic trends produced by GANs, SDN controllers can optimize the allocation of network resources, reducing latency and enhancing QoS.RNN-based authorization rules continuously identify trends in network activity, minimising potential vulnerabilities and adapting to new threats.
- By automating both access control and traffic design, the method lessens the operational load on the network's management and promotes more effective resource utilization.
- Assessing the efficacy of the hybrid GAN-RNN models in SDN scenarios through multiple simulations and real-world experiments. The results show notable gains in network security and effectiveness, highlighting the approach's potential for contemporary network management.
- This study offers a ground-breaking framework for utilizing the powers of hybrid GAN-RNN models to optimize software-defined networking. The next phase of network administration is anticipated with the

integration of flexible access controls and realistic traffic generation.

The Section I provides an overview of the paper. The Section II reviews existing literature and emphasizes the gap in addressing techniques for network security enhancement. Section III defines the central research problem concerning driver drowsiness detection complexities. Section IV outlines data collection, preprocessing, feature extraction, and the integration of Hybrid GAN-RNN. Section V presents empirical findings, compares classifier performance, and explores implications and future research directions is in Section VI, which is solidifying the research's significance in Network security.

II. Related Works

Ramprasath and Seethalakshmi [11] examines a crucial element of SDN, focusing on the requirement for improved security controls in SDN systems. By separating the information plane from the control plane, SDN enables on-demand services and the ability to configure networks dynamically. The study correctly highlights the fact that, despite the fact that SDN controls traffic flows and flow labels based on Open Flow virtual switches successfully, it lacks built-in security safeguards to combat malicious traffic, such as Denial-of-Service (DoS) assaults, which can significantly lower service availability. It is laudable that the paper's main emphasis is on identifying and reducing DoS threats by dynamically setting firewalls within SDN setups. The study makes an effort to close this security gap by using dynamic access control lists. This study's noteworthy feature is the use of Mininet to simulate SDN with dynamic access control list attributes. This enables real-world testing and experimentation. The practical relevance of the findings is given more weight by this empirical confirmation. The work would benefit from a more thorough examination of the exact methods and tools employed for DoS attack detection and mitigation inside the SDN environment to further strengthen its contribution. Readers would have a better grasp of the suggested strategy if more information was provided about the dynamic access control list implementation and the standards for differentiating malicious from normal traffic. In order to reduce DoS attacks, the article tackles a critical security issue in SDN systems and offers a potential solution using dynamic access control lists. The research offers important insights towards strengthening the security of SDN systems by fusing theoretical understanding with real-world testing. The study would be even more helpful and effective if it elaborated on the technical specifics of the suggested strategy.

Vimal et al. [12] provides a fascinating and current investigation into how integrating Internet of Things (IoT) devices might improve the security of Software-Defined Networks (SDN), with an emphasis on boosting information access control using encryption. The research's emphasis on creating a strong infrastructure for IoT devices is well-placed given the quick spread of IoT devices. The notion of a stability routing protocol, which evaluates the reliability of devices and packet flows, is introduced in this work. To create dependable SDN routes, this method makes use of the mutual trust between network components, Quality of Service (QoS), and energy circumstances. An important advancement is the incorporation of SDN architecture into the Cognitive Protocol Network (CPN) technology platform to improve energy efficiency. A novel strategy for tackling security issues is the use of stochastic neural networks (SNNs) for decentralized decision-making based on data gleaned from perceptual packets. It is a praiseworthy effort to include these components into SerIoT approaches to provide IoT encryption for information access control. The implementation of various techniques and technologies, particularly the precise approaches utilized for IoT encryption and access control, might need more explicit explanations in the study. The complexity of the study would also be increased by providing more detail on how the suggested network infrastructure solves issues like erratic connectivity, constrained cryptographic capacity, and energy restrictions. The study emphasizes the significance of tackling cluster instability for platform efficiency as well as the necessity of collaboration. A deeper grasp of the research's practical relevance might be provided by providing additional information on the difficulties and potential solutions linked to these issues.

Shin et al. [13] addresses the potential of Software-Defined Networking to improve network security as it goes into a crucial and modern junction of technology. Because it can separate control logic from conventional network hardware, SDN has attracted a lot of interest as a transformational technology that can improve network administration and innovation. The authors note that despite its capabilities, SDN is still largely disregarded by the security community, highlighting the underutilized potential of SDN in the area of network security. The article presents a thorough overview of the prospects offered by this technology by meticulously evaluating how the distinctive features and capabilities of SDN may strengthen network security and the larger information security process. This in-depth analysis of SDN's potential to advance network security research creates fresh directions for future study in this crucial area. The article does a good job of outlining the main ideas and goals, but it might make a bigger impact if it went into more detail with examples or case studies of how SDN has been used to successfully handle security issues. Giving readers specific examples of how SDN is used to enhance network security can help readers understand the real-world applications of the technology and may encourage more research projects. The report effectively discusses the important role that SDN may play in strengthening network security and sheds light on an exciting yet underappreciated field of study. It is a useful tool for academics and professionals who want to strengthen network security in the context of a changing technological environment by utilizing SDN's capabilities. Extending on actual use cases and useful implementations might increase the paper's impact and usefulness.

Ahmad et al. [14] focuses on the use of machine learning (ML) approaches to thwart Denial of Service (DoS) and Distributed DoS (DDoS) attacks inside the SDN framework. It provides a timely analysis of the essential confluence between Software Defined Networking (SDN) and security. With its logically centralized control plane, SDN offers enhanced network administration as a viable response to a number of issues in conventional networks. But because of the security flaws introduced by this centralization, SDN control systems are becoming tempting targets for malicious attacks. Given ML's shown efficacy in finding security vulnerabilities, the paper made a sound decision to use ML approaches for recognizing and mitigating DoS and DDoS attacks. It is a useful addition to test these ML approaches in practice in an SDN system, especially by subjecting the SDN controller to DDoS attacks. It offers useful perceptions on the applicability and constraints of ML-based security methods for upcoming communication networks. The work may benefit from a more in-depth examination of the various ML approaches used and the standards for judging their efficacy. Readers would comprehend the use of ML models or algorithms to SDN security more clearly if examples or case studies of these applications were given. This article discusses security flaws resulting from centralized control, a serious issue in the SDN space. The work offers a significant addition to the area by outlining and assessing ML strategies to defend against DoS and DDoS assaults within SDN. It highlights the value of ML-based solutions in securing upcoming communication networks and provides a viable path for boosting network security. The paper's usefulness and effect would be increased with additional clarification of the ML approaches employed.

Pérez-Díaz et al. [15] provides a significant and pertinent addition to the continuing problem of LR-DDoS attack mitigation in the context of Software-Defined Networks (SDN). Due to the notoriously difficult-to-detect nature of LR-DDoS assaults and the potential harm they pose in SDN environments, a flexible modular architecture for their detection and mitigation has been developed. The study utilizes Machine Learning models such as J48, Random Tree, REP Tree, Random Forest, Multi-Layer Perceptron and Support Vector Machines to train an Intrusion Detection System (IDS). Despite the inherent difficulties presented by LR-DDoS assaults, the assessment of these ML models using the Canadian Institute of Cyber security (CIC) DoS dataset showed a remarkable detection rate of 95%. One of the key advantages of this study is the practical implementation of the open network operating system (ONOS) controller within a Mini-net virtual machine, which tries to faithfully imitate real production network circumstances. This strategy strengthens the paper's credibility and explains how it may be used in real-world network security settings. The article also emphasizes that the intrusion prevention detection system successfully mitigates all assaults identified by the IDS, highlighting the usefulness of the suggested architecture in LR-DDoS attack detection and mitigation. This study presents a novel approach that combines ML approaches with a flexible modular architecture to solve the ongoing problem of LR-DDoS assaults in SDN systems. Its potential as a formidable tool for network security is highlighted by its easy deployment and remarkable detection results. The impact of the work would be increased and new insights would be provided for security practitioners and academics with a more thorough investigation of the difficulties and model choices.

Latif et al. [16] discusses security, a key concern in the context of the Industrial Internet of Things environment. Smart cities, agriculture, and healthcare are just a few areas where IIoT is crucial due to its integration of sensors, devices, and databases. This study acknowledges the distinct security risks that the IIoT presents as a result of its integration into more complex operational systems. In this research, a unique method for anticipating and detecting several cybersecurity attacks—including denial of service, malicious operation, malicious control, data type probing, espionage, scan, and incorrect setup—that are frequently seen in IIoT contexts is presented. It undertakes a comparison analysis with conventional machine learning methods including artificial neural networks, support vector machines, and decision trees, and proposes a lightweight random neural network (RaNN) as the foundation for its prediction model. The study's main conclusions show that the suggested RaNN-based model performs admirably, with accuracy rates of 99.20%, precision, recall, and the F1 score all above 99%. The model also has short prediction duration of 34.51 milliseconds. These findings show how well the RaNN model predicts and recognises IIoT cybersecurity threats. The work makes an important addition to the area since it tackles the urgent demand for reliable security solutions in IIoT. It is an intriguing option for boosting IIoT security since it uses a lightweight RaNN model and performs better than conventional approaches.

The claimed accuracy gains of 5.65% for IoT security over cutting-edge machine learning algorithms is notable and demonstrates the practical applicability of this study. Nevertheless, when using this approach in complex IIoT contexts, it's critical to take into account potential limits, such as the range and variety of attack scenarios, and scalability, including real-world deployment issues. Despite this, the article offers a useful framework for more study and advancement in the field of IIoT security, including the potential for real-world use in securing vital industrial systems. While the previously discussed works provide important insights into various aspects of protecting SDN and addressing cybersecurity concerns within the IIoT framework, a common shortcoming of these studies is the lack of comprehensive real-world implementation and evaluation. Although some studies employ simulation techniques, little is known regarding whether the proposed solutions can be scaled and applied in complex, real-world large-scale network environments. A more thorough analysis of the potential challenges and setbacks that came across throughout the development of their safety processes, such as issues with interoperability, resource constraints, and adaptability to evolving attack strategies, would also greatly increase the study's practical significance and utility.

III. Problem Statement

Although SDN presents the possibility of dynamic and adaptable network management, efficiency optimization and security remain major obstacles. Lack of integrated security measures to thwart malicious traffic, particularly Denial-of-Service (DoS) attacks, is one of the major problems that can seriously impair service availability. Many times, existing SDN solutions are unable to adequately handle these security issues. As such, the development of a comprehensive strategy that enhances security protocols while simultaneously optimizing network performance is imperative. The purpose of this research is to determine how well a hybrid GAN-RNN approach, which sets firewalls dynamically using dynamic access control lists, can handle this dual challenge. Furthermore, it aims to provide a new solution that improves network safety and efficiency in SDN environments.

The ability of the proposed Hybrid GAN-RNN method to focus on both security concerns and network efficiency optimizing in SDN systems is what makes it effective. By using Generative Adversarial Networks (GANs) to simulate feasible traffic patterns and Recurrent Neural Networks (RNNs) for adaptable controls on access, this method offers a multidimensional solution. GANs can be used to model various traffic scenarios, and RNNs can be used to flexibly set access control rules based on real-time threat detection to achieve optimal network designs. This hybrid paradigm effectively adapts to changing network conditions and security threats, significantly enhancing the overall resilience of SDN systems. Moreover, employing the Mini net for real-world assessment boosts the practical value of the results and increases the possibility of their successful implementation in operational networks. More information about the specific methods and tools employed for DoS attack detection and avoidance in an SDN environment is required in order to improve the strategy's effectiveness.

IV. Proposed Hybrid GAN-RNN for Network Security

The proposed methodology represented in Fig. 1 starts with the collection of network traffic data, which is then meticulously pre-processed to cleanse, format, and extract relevant features. A tailored Generative Adversarial Network (GAN) architecture is crafted to generate synthetic traffic patterns that closely resemble real network data. These synthetic patterns are crucial for enhancing security analysis. Simultaneously, a Recurrent Neural Network (RNN) is employed to predict network attacks based on the generated traffic patterns. The RNN learns to recognize temporal patterns and anomalies in the data, aiding in the proactive identification of potential security threats. Following the hybrid GAN-RNN approach, the system's performance is thoroughly analyzed, assessing its ability to generate realistic traffic and predict attacks accurately. Additionally, a comparative evaluation is conducted to benchmark the proposed methodology against existing approaches, providing insights into its effectiveness in bolstering network security.

A. Data Collection

The CICIDS2017 dataset provides a valuable resource for improving network performance and security through the utilization of Hybrid GAN-RNN models within Software-Defined Networking (SDN) environments. This dataset incorporates both benign network traffic and a wide range of common attacks, making it a suitable foundation for our research and development in the field of networks security and optimization. The CICIDS2017 dataset offers a comprehensive view of network traffic, encompassing benign background traffic and real-world attack scenarios. The dataset is constructed with meticulous attention to realism, featuring the following key components: Generated using the B-Profile system, the dataset simulates the naturalistic behaviors of 25 users engaging in various protocols such as email, HTTP, FTP, HTTPS, and SSH. This component mimics real-world user interactions, contributing to the authenticity of the dataset. The dataset represents a complete network infrastructure, including components like Modem, Firewall, Switches, Routers, and a diverse array of operating systems (e.g., Ubuntu, Windows, and Mac OS X). This realistic topology ensures that the dataset mirrors complex network environments. The CICIDS2017 dataset incorporates real attacks from the Attack-Network, enabling researchers to analyze and develop security measures against a wide range of threats. This includes the most up-to-date common attacks, adding relevance to the research context [17].

B. Data Pre-processing using Handling Missing Values

The time series dataset representing network-wide traffic states, which is denoted as$X$. This dataset encompasses observations collected over time, each of which corresponds to a specific time step. The dimensionality of the dataset is determined by the amount of sensor stations in the network, denoted as$D$. Mathematical representation of this time series dataset X as follows in Eq. (1):

$$X = \{x_1, x_2, \dots, x^{\mathrm{T}}\}^T \in \mathbb{R}^{T \times D} \quad (1)$$

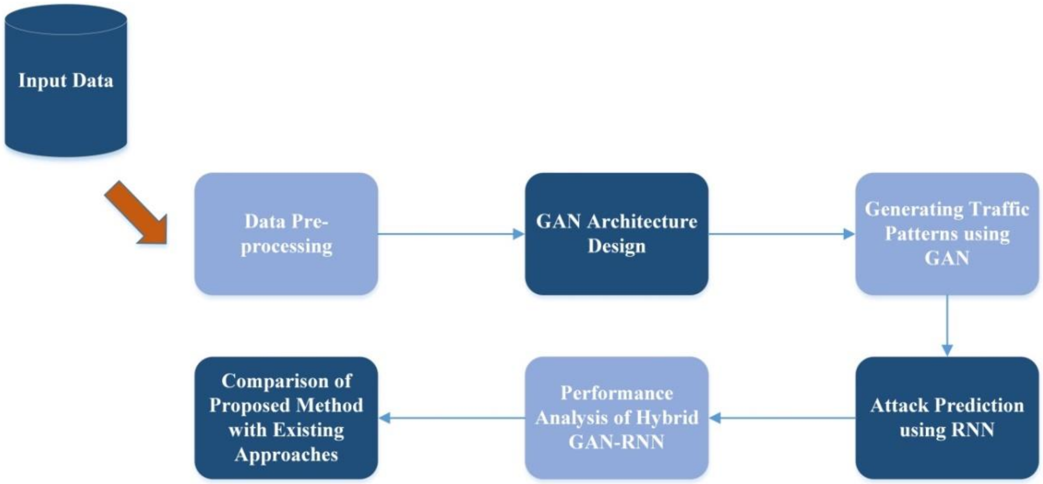

Input Data
Data Pre-processing
GAN Architecture Design
Generating Traffic Patterns using GAN
Attack Prediction using RNN
Performance Analysis of Hybrid GAN-RNN
Comparison of Proposed Method with Existing Approaches

Fig. 1. Proposed workflow.

T Signifies the total number of time steps in our dataset. D signifies the number of sensor stations distributed across the network. Each vectorx^t, associated with time step$t$, belongs to the real numbers$\mathbb{R}^D$. This vector encapsulates the traffic state information of the D sensor stations at that specific time [18].

Within this vector, each element x_t^d corresponds to the traffic speed observed at the d-th sensor station. This research discusses "traffic state," It specifically focuses on traffic speed. This definition aligns with the characteristics of the datasets we employ, particularly those used in our experimental section. Traffic sensors, such as inductive looping detector, may encounter failures due to various reasons, including wire insulations breakdown, damage from building activities, or electronic unit failures. These sensors failures result in missing values within our collected data. To address the issue of missing values, Research employs a masking vectormt, which is binary and takes values from the set {0, 1}, to indicate whether traffic states are missing at a specific time step t. The masking vectors for xt is defined as follows in Eq. (2):

$$m_t^d = \begin{cases} 1, if\ x_t^d\ \ is\ observed \\ \quad 0, otherwise \end{cases} \qquad (2)$$

Consequently, for a given traffic state data sample X in$\mathbb{R}^{T\times D}$, derivation of a corresponding masking data sample M, is represented in Eq. (3):

$$M = \{m_1, m_2, \dots, m^T\} \in \mathbb{R}^T \times D \qquad (3)$$

The traffic state prediction problem revolves around the objective of learning a function$F$ $(\cdot)$. This function is designed to map T historical traffic state datas observations to the subsequent traffic state data at the next time step. This problem can be formally described as in Eq. (4):

$$F([x_1, x_2, \dots, x^T];\ [m_1, m_2, \dots, m^T]) = [x_{T+1}] \qquad (4)$$

where, F aims to predict the traffic state at time step $T + 1$ depend on the historical traffic state data up to time step$T$, taking into account the masking information to handle missing values.

C. Hybrid GAN-RNN Architecture for Generating Traffic Patterns and Access Control

The Hybrid GAN-RNN architecture is designed to enhance network security by generating realistic network traffic patterns and making access control decisions based on those patterns. This architecture comprises two main components: a Generative Adversarial Network (GAN) for traffic pattern generation and a Recurrent Neural Network (RNN) for access control. The hybrid GAN-RNN architecture is shown in Fig. 2. In the GAN component, the generator (G) takes random noise (z) as input and generates synthetic traffic patterns (X_synthetic). The discriminator (D) then evaluates these synthetic patterns and real traffic patterns (X_real), aiming to distinguish between them. The objective is to train the generator to produce traffic patterns that are indistinguishable from real ones, while the discriminator becomes more adept at differentiating real from synthetic patterns. This adversarial training process is guided by a GAN loss function that encourages the generator to improve its pattern generation capabilities. The discriminator's formal objective is to acquire characteristics θd that maximize the likelihood of properly categorizing both training and produced data; the generator's objective is to discover settings θg that minimize $log\ 1 - D(G(z))$. the following two-player minimax game with value functions $V(G, D)$is therefore played by the two neural networks.

$$V(G, D) = max\ D\ min\ G\ \{Ex\ [log\ D(x)] + Ez[log(1 - D(G(z)))]\} \qquad (5)$$

where, G(z) is the created false data provided by the noise vector z, D(G(z)) is the estimated chance of a fake instance being honest, and D(x) is the estimated likelihood of an actual model being real generated by the discriminating neural networks. The generator theoretically learns to produce genuine samples when it reaches equilibrium.

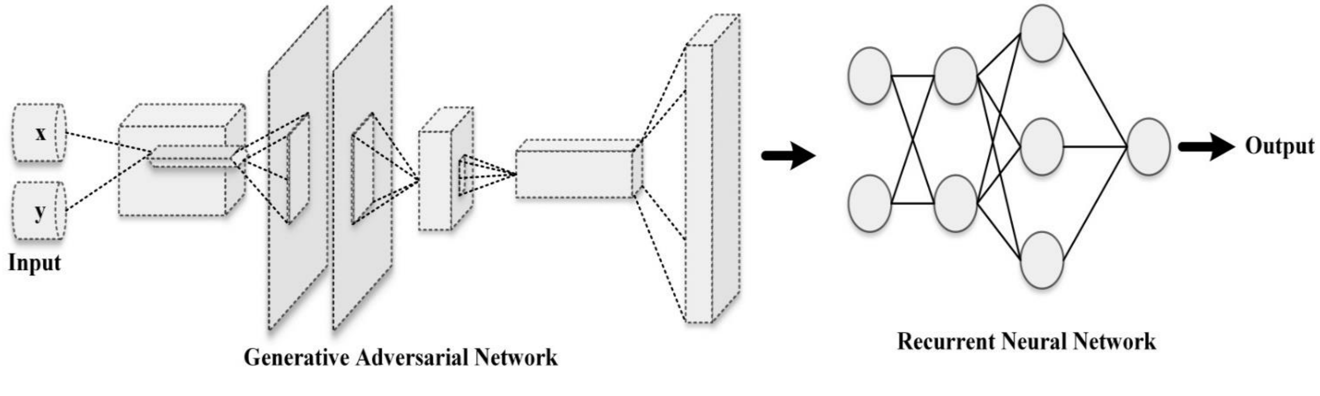


Fig. 2. Hybrid GAN-RNN architecture.

```
Algorithm 1: Hybrid GAN-RNN Algorithm for Network Security Enhancement
Input:  Network Traffic data
Output: Predicting the attack in Network traffic data
Load input data
                    $X = \{x_1, x_2, \ldots, x^{\mathrm{T}}\}^T \in \mathbb{R}^{T\times D}$          // data acquisition
Preprocess network traffic data
    Cleanse and Handling missing traffic data's, and normalize.                          //handling missing values
Split the data into training and testing sets.
Generation of Traffic Patterns                                                           // GAN Training
    Initialize the GAN model with a generator (G) and discriminator (D).
    Train the GAN by iteratively optimizing G and D
    Generate synthetic traffic patterns using G
    Calculate the GAN loss based on D's ability to distinguish real from synthetic patterns.
    Back propagate the loss and update the G and D weights.
    Repeat until convergence or a predefined number of epochs
Attack Prediction                                                                        //RNN Training
    Initialize the RNN model for attack prediction.
    Define the RNN architecture, loss function, and optimizer.
    Train the RNN using the generated synthetic traffic patterns
    Input the synthetic traffic data sequence to the RNN
    Calculate the loss based on the predicted attacks and actual labels (ground truth).
    Back propagate the loss and update the RNN weights
    Repeat until convergence or a predefined number of epochs
 Prediction of Attack in Network                                                         //RNN
Evaluate the hybrid GAN-RNN system's performance using testing data                      //Performance Evaluation
    Measure the accuracy of attack predictions
    Calculate other relevant metrics such as precision, recall, and F1-score
    Assess the quality of generated traffic patterns
```

The RNN component processes the generated traffic patterns (X_synthetic) and performs access control. For this purpose, the RNN can employ a LSTM architecture, which allows it to consider temporal dependencies in the traffic data. The RNN's internal state (ht) evolves as it processes the traffic patterns, and at each time step, it produces access control decisions (yt) through a Softmax layer. These decisions can take various forms, such as binary access control (allows or deny) or multiclass access policies based on the traffic content and context.

The key innovation of this architecture lies in its combination of GAN and RNN components. The GAN generates synthetic traffic patterns that are realistic and diverse, reflecting various network activities. The RNN, in turn, leverages these patterns to make access control decisions in real-time. This approach enables a more dynamic and adaptable access control system that can respond effectively to evolving network conditions and potential security threats. During training, the entire hybrid architecture is optimized through a joint loss function that balances the GAN loss and the access control loss. This ensures that the generated traffic patterns are not only realistic but also suitable for access control decision making. The RNN's parameters are fine-tuned to make accurate access control decisions based on the generated patterns, thereby enhancing network security.

V. Results and Discussion

The result section provides a comprehensive evaluation of the proposed network security enhancement method, employing various evaluation metrics such as accuracy, precision, recall, and F1-score. The analysis begins with a comparison of the method's accuracy on different datasets, highlighting the notably high accuracy of the "Proposed GAN-RNN" approach on the CICIDS2017 dataset. A comparative assessment with existing methods further underscores the method's superiority, showcasing exceptional precision, recall, and F1-score. Graphs depict the performance trends, demonstrating the model's convergence and its ability to generalize to unseen information. The training and testing graphs illustrate the model's progression, while the loss graph reveals its capacity to avoid overfitting. The results validate the effectiveness of the proposed methods in network intrusion detection, emphasizing its potential to enhance network security with impressive accuracy and robustness. The practical usefulness of the Hybrid GAN-RNN technique extends to dynamic business networks, allowing for adaptive routing and resource allocation. Its 99.4% accuracy in cybersecurity guarantees quick access policy changes, which are essential for sectors like banking and healthcare and improve overall network security and efficiency.

A. Evaluation Metrics

Four assessment measures were used in the study to evaluate the designs: F1-score, accuracy, precision, and recall. Such specific variables are described as in Eq. (6), (7), (8) and (9):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$F1score = \frac{2*Recall*Precision}{Recall+precision} \tag{9}$$

TP is the number of information that, irrespective of every one of the types of information which was genuinely positive, was precisely identified as positive. TN is the number of information that, irrespective of all the results which was truly negative, were properly identified as negatives. The number of variables that the equation incorrectly categorized as negative despite the fact they had been positive in the input data is represented by the letter FN. The number of values that the algorithm incorrectly categorized as positive when they had been negative in the source data is known as false positives, or

FP. The percentage of the number of information that the algorithm identified as being positive to the number of positive results that were really present in the collection of data is known as recall. Precision can be defined as the proportion of the entire amount of information that the model properly identified as positive to the number of data which the algorithm categorized as positive. Lastly, as mentioned in, the F1-score represents the harmonious average of recall and precision Top of Form [19].

This Table I present the performance results of different intrusion detection methods on these datasets, with the "Proposed GAN-RNN" method achieving notably high accuracy on the CICIDS2017 dataset at 99.4%. It's important to note that the choice of dataset and the specific evaluation metrics used can significantly impact the reported accuracy, and the effectiveness of a method may vary depending on the dataset's characteristics and the complexity of the network security task.

The graph in Fig. 3 depicts, the methods consistently outperform others across multiple datasets and which ones may excel in specific contexts. This comparison aids in the selection of the most robust intrusion detection method for diverse network security environments, contributing to informed decision-making in network security strategy.

Table II presents a comparative overview of different methods applied to network intrusion detection, showcasing their performance across multiple evaluation metrics. It includes the methods GRU, CNN, B-GRU, and the "Proposed GAN-RNN." the "Proposed GAN-RNN" method exhibits exceptional accuracy, achieving an impressive 99.4%, surpassing the other methods in the accuracy metric.

Furthermore, it excels in precision with a score of 99.25%, ensuring a high proportion of correctly classified positive predictions. It demonstrates remarkable recall at 99.6%, effectively capturing a significant portion of actual positive instances. The F1-score, a balanced measure of precision and recall, remains strong at 99.4%, further affirming the method's robustness in network intrusion detection, making it a highly promising approach for bolstering network security.

The Graph represents in Fig. 4 shows the performance comparison of different intrusion detection methods on various metrics, including accuracy, precision, recall, and F1-score.

Fig. 5 represents the training and testing graph for the proposed network security enhancement method illustrates the model's performance throughout the training process. During the training phase, the metrics are plotted as they evolve with each epoch, showing how the model learns and improves its performance over time. The testing phase is also depicted on the same graph, showcasing how the model generalizes to unseen data. This graph provides a clear visualization of the model's convergence and its ability to avoid overfitting or underfitting, thus assisting in the evaluation and refinement of our network security enhancement approach. The testing accuracy attained is 99.4%.

Fig. 6 represents the training and testing loss graph is a graphical representation that illustrates the changes in the loss function values of a machine learning or deep learning model during both the training and testing phases. The testing loss curve reveals the model generalizes to unseen data, and ideally, it should exhibit a similar decreasing trend, indicating that the model is not overfitting.

TABLE I. ACCURACY COMPARISON OF DATASET

Dataset	Methods	Accuracy
KDD99 [20]	DT	92.3
UNSW-NB15 [20]	LR	85.56
CICIDS2017	Proposed GAN-RNN	99.4

TABLE II. PERFORMANCE COMPARISON WITH EXISTING METHODS

Methods	Accuracy	Precision	Recall	F1-score
GRU [21]	99	99	99	99
CNN [21]	97.7	97.4	98.2	99
B-GRU [21]	98.74	98.9	99	99
Proposed GAN-RNN	99.4	99.25	99.6	99.4

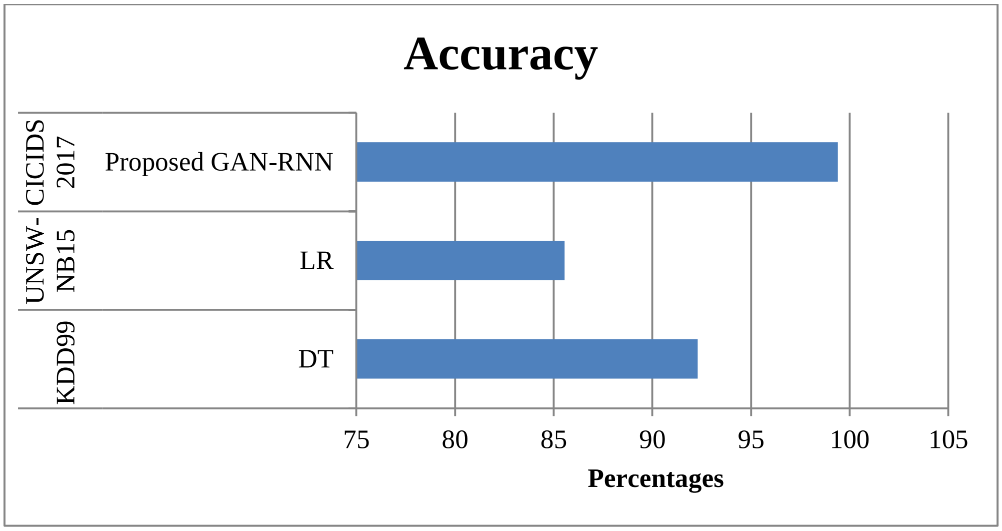


Fig. 3. Dataset comparison.

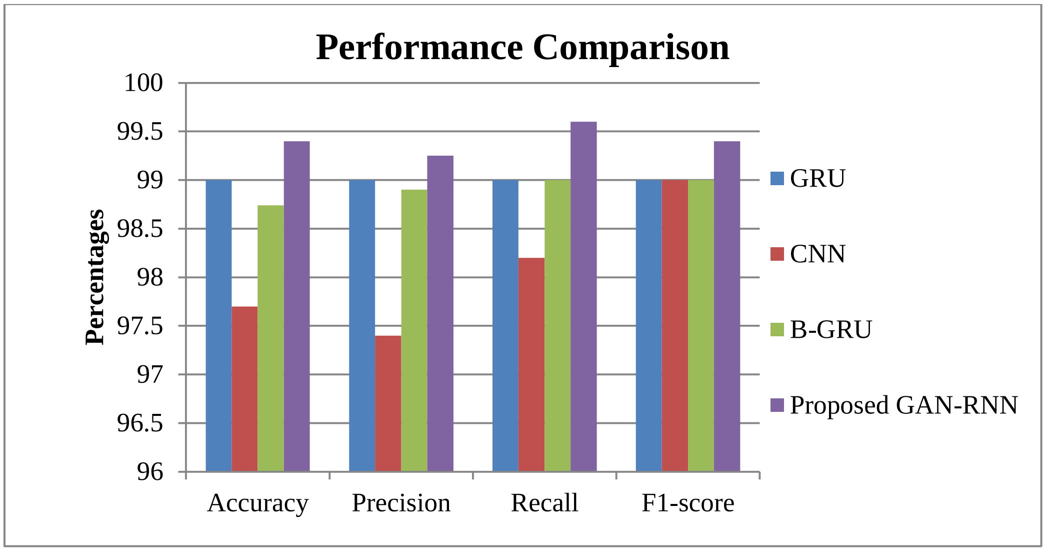


Fig. 4. Evaluation of performance with existing approaches.

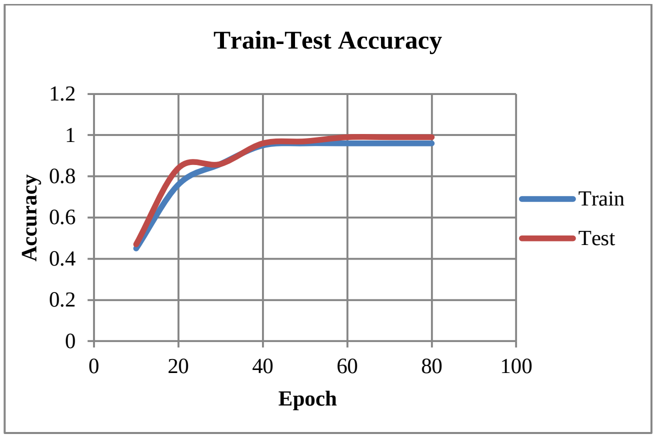


Fig. 5. Training and testing accuracy.

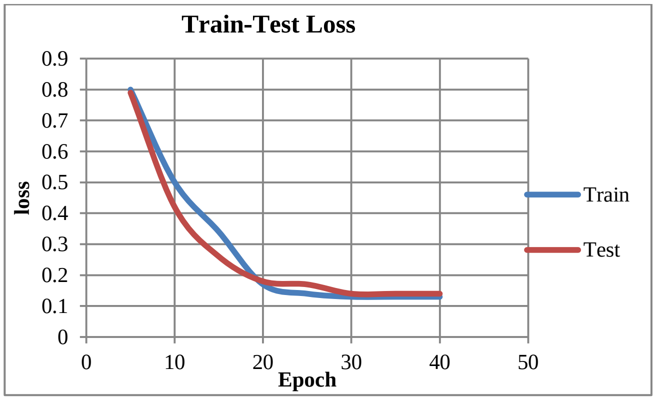


Fig. 6. Training and testing loss.

The Fig. 7 displays a Generative Adversarial Network with Recurrent Neural Network (GAN-RNN) model's Receiver Operating Characteristic (ROC) curve. The model's ability to distinguish between classes, especially in issues with binary classification, is represented graphically by the ROC curve. The true positive rate (sensitivity) during each threshold, which ranges from 0 to 0.7, is paired with a corresponding threshold value in the table. The true positive rate tends to rise in tandem with the threshold, suggesting that the model is becoming more accurate at identifying positive instances. The true positive rate increases gradually from 0.07 to 0.994 to be the threshold increases, indicating that the GAN-RNN model has high discriminatory power. This indicates that the model performs well in classifying positive instances, demonstrating its efficacy in achieving high sensitivity across a range of threshold values.

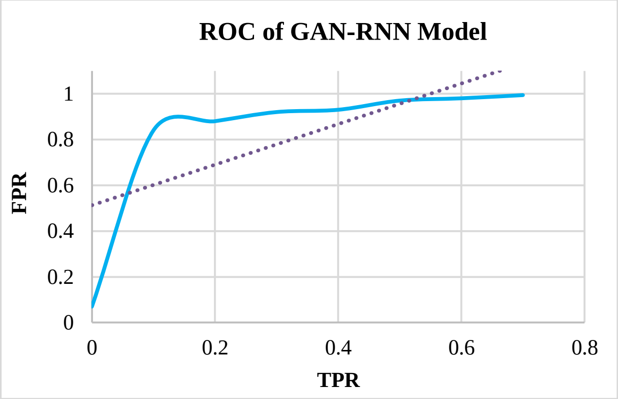


Fig. 7. ROC of GAN-RNN model.

B. Discussion

The accuracy of several methods for identifying intrusions on a range of datasets is assessed at the outset of the findings section. The efficiency results for the three different datasets—KDD99, UNSW-NB15, and CICIDS2017—are shown in Table I. On the CICIDS2017 dataset, the "Proposed GAN-RNN" approach notably obtains a very high accuracy value of 99.4%. This extraordinary accuracy demonstrates how well the approach works in a specific dataset to recognize network security problems. However, given that network data can differ greatly in complexity and feature sets, it is imperative to recognize that the selection of dataset is a critical factor in deciding accuracy. As a tool for comparing datasets, the chart in Fig. 3 shows how the techniques continually perform better than others on a variety of datasets. It makes it possible to pinpoint the approaches that work best in certain situations. The comparison is essential because it helps choose the most reliable intrusion detection technique for various network security scenarios. The graph presents the overall efficacy of the "Proposed GAN-RNN" approach, indicating a potential option for improving network security on a variety of datasets.

Table II provides an extensive comparative analysis of various intrusion detection techniques, such as GRU, CNN, B-GRU, and the "Proposed GAN-RNN." The "Proposed GAN-RNN" approach performs exceptionally well according to a number of evaluation parameters. It is noteworthy for achieving a 99.4% accuracy rate, which is higher than the other approaches. The technique also performs exceptionally well in terms of precision (99.25%), guaranteeing a large percentage of accurately identified positive predictions. Furthermore, it exhibits exceptional memory (99.6%), successfully catching a substantial proportion of true positive cases. At 99.4%, the F1-score—a measure that strikes a compromise between recall and precision—remains robust. All of these findings support the "Proposed GAN-RNN" approach's robustness and dependability in detecting network intrusions. Because of its exceptional performance, the approach offers both accuracy and precision in spotting security issues, making it a very viable alternative to strengthen network security. The suggested network security enhancement approach is thoroughly evaluated in the results section, which highlights its remarkable accuracy, precision, recall, and F1-score. It demonstrates how the approach can continuously beat other approaches on various datasets, which makes it a strong option for secure network application. These results provide insightful information that may be used to make well-informed decisions about network security planning and technological implementation. The access control as well as traffic engineering systems that are now dependent on network security may not be scalable, may have trouble adapting to changing threats in real-time, and may find it difficult to properly handle growing cyber threats [8].

VI. Conclusion and Future Work

In the framework of SDN-based traffic management and access control, hybrid GAN-RNN models were proposed and their efficacy was shown in the present investigation. The results obtained suggest that this novel technique holds significant potential for improving software-defined network security and performance. The Hybrid GAN-RNN design has demonstrated notable gains in network effectiveness and threat reduction through the creation of realistic patterns of traffic and accurate access control choices. In today's intricate and constantly evolving network systems, the capacity to detect abnormalities, adjust to changing conditions, and optimize traffic flows is an essential skill. The suggested technique's high recall, accuracy, and precision highlight its potential as a vital resource for network managers and security experts. This strategy has the ability to enable enterprises to strengthen their safety posture, maximize resource efficiency, and handle their networks more effectively as the network environment changes. Through the integration of Hybrid GAN-RNN models within SDN, this study considerably improves knowledge while improving the efficiency and security of networks. Results validate hypotheses and lay the groundwork for more investigation into adaptive access management and optimizing traffic in dynamic contexts in the future.

Provide means by which access control policies can be automatically modified in response to threat assessments and network conditions in real time, enabling more flexible and responsive security. Instead of depending only on historical data, investigate real-time analysis abilities that allow the system to identify and address security threats and operational issues as they arise. To convert created traffic trends into useful network configurations and rules, tighten the connection with SDN controllers. In multi-domain or mixed-cloud situations, when network intricacy and safety issues are heightened, expand the Hybrid GAN-RNN technique to optimize network security and performance. To ensure resilience in the midst of sophisticated dangers, assess the proposed method's resistance against adversarial assaults that attempt to interfere with traffic patterns or evade control of access. To handle increasing threats, future research might focus on improving the hybrid GAN-RNN method's scalability and flexibility. A more thorough and forward-thinking approach would involve looking into effective ways of managing larger networks and new security challenges.

References

[1] P. Mishra, V. Varadharajan, U. Tupakula, and E. S. Pilli, “A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection,” IEEE Commun. Surv. Tutorials, vol. 21, no. 1, pp. 686–728, 2019, doi: 10.1109/COMST.2018.2847722.

[2] V. Kapoor and R. Yadav, “A Hybrid Cryptography Technique for Improving Network Security,” IJCA, vol. 141, no. 11, pp. 25–30, May 2016, doi: 10.5120/ijca2016909863.

[3] C. Yu, J. Lan, Z. Guo, and Y. Hu, “DROM: Optimizing the Routing in Software-Defined Networks With Deep Reinforcement Learning,” IEEE Access, vol. 6, pp. 64533–64539, 2018, doi: 10.1109/ACCESS.2018.2877686.

[4] N. Awadallah Awad, “Enhancing Network Intrusion Detection Model Using Machine Learning Algorithms,” Computers, Materials & Continua, vol. 67, no. 1, pp. 979–990, 2021, doi: 10.32604/cmc.2021.014307.

[5] Muthukumaran V., V. V. Kumar, R. B. Joseph, M. Munirathanam, and B. Jeyakumar, “Improving Network Security Based on Trust-Aware Routing Protocols Using Long Short-Term Memory-Queuing Segment-Routing Algorithms:,” International Journal of Information Technology Project Management, vol. 12, no. 4, pp. 47–60, Oct. 2021, doi: 10.4018/IJITPM.2021100105.

[6] S. Akbar, J. A. Chandulal, K. N. Rao, and G. S. Kumar, “Improving network security using machine learning techniques,” in 2012 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India: IEEE, Dec. 2012, pp. 1–5. doi: 10.1109/ICCIC.2012.6510197.

[7] R. Ahmad, R. Wazirali, and T. Abu-Ain, “Machine Learning for Wireless Sensor Networks Security: An Overview of Challenges and Issues,” Sensors, vol. 22, no. 13, p. 4730, Jun. 2022, doi: 10.3390/s22134730.

[8] S. Anbalagan et al., “Machine-Learning-Based Efficient and Secure RSU Placement Mechanism for Software-Defined-IoV,” IEEE Internet Things J., vol. 8, no. 18, pp. 13950–13957, Sep. 2021, doi: 10.1109/JIOT.2021.3069642.

[9] S. Nanda, F. Zafari, C. DeCusatis, E. Wedaa, and B. Yang, “Predicting network attack patterns in SDN using machine learning approach,” in 2016 IEEE Conference on Network Function Virtualization and

Software Defined Networks (NFV-SDN), Palo Alto, CA: IEEE, Nov. 2016, pp. 167–172. doi: 10.1109/NFV-SDN.2016.7919493.

[10] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications," IEEE Commun. Surv. Tutorials, vol. 16, no. 4, pp. 1996–2018, 2014, doi: 10.1109/COMST.2014.2320099.

[11] J. Ramprasath and V. Seethalakshmi, "Secure access of resources in software-defined networks using dynamic access control list," Int J Communication, vol. 34, no. 1, p. e4607, Jan. 2021, doi: 10.1002/dac.4607.

[12] V. Vimal et al., "Enhance Software-Defined Network Security with IoT for Strengthen the Encryption of Information Access Control," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–10, Oct. 2022, doi: 10.1155/2022/4437507.

[13] S. Shin, L. Xu, S. Hong, and G. Gu, "Enhancing Network Security through Software Defined Networking (SDN)".

[14] A. Ahmad, E. Harjula, M. Ylianttila, and I. Ahmad, "Evaluation of Machine Learning Techniques for Security in SDN," in 2020 IEEE Globecom Workshops (GC Wkshps, Taipei, Taiwan: IEEE, 2020, pp. 1–6. doi: 10.1109/GCWkshps50303.2020.9367477.

[15] J. A. Pérez-Díaz, I. A. Valdovinos, K.-K. R. Choo, and D. Zhu, "A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning," IEEE Access, vol. 8, pp. 155859–155872, 2020, doi: 10.1109/ACCESS.2020.3019330.

[16] S. Latif, Z. Zou, Z. Idrees, and J. Ahmad, "A Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network," IEEE Access, vol. 8, pp. 89337–89350, 2020, doi: 10.1109/ACCESS.2020.2994079.

[17] "CICIDS2017." https://www.kaggle.com/datasets/cicdataset/cicids2017 (accessed Sep. 18, 2023).

[18] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Forecasting Network-wide Traffic State with Missing Values." arXiv, May 23, 2020. Accessed: Sep. 19, 2023. [Online]. Available: http://arxiv.org/abs/2005.11627.

[19] B. Jang, M. Kim, G. Harerimana, S. Kang, and J. W. Kim, "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism," Applied Sciences, vol. 10, no. 17, p. 5841, Aug. 2020, doi: 10.3390/app10175841.

[20] N. Moustafa and J. Slay, "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," Information Security Journal: A Global Perspective, vol. 25, no. 1–3, pp. 18–31, Apr. 2016, doi: 10.1080/19393555.2015.1125974.

[21] H. Wang and W. Li, "DDosTC: A Transformer-Based Network Attack Detection Hybrid Mechanism in SDN," Sensors, vol. 21, no. 15, p. 5047, Jul. 2021, doi: 10.3390/s21155047.

The Contribution of Health Management Information Systems to Enhancing Healthcare Operations

Majzoob K.Omer
Department of Computer Science, Al-Baha University, Al-Baha, Saudi Arabia

***Abstract*—Various strategies for enhancing quality have been implemented by developed and developing countries in light of the worldwide emphasis on bolstering healthcare systems. Many nations are currently directing their attention towards bolstering their existing information systems or establishing new ones, recognizing the critical role of information in the functioning of healthcare systems. The study aimed to assess the impact of leadership style, organizational factors, technology, and healthcare provider behavior on the implementation of health management information systems in healthcare organizations. While the study was informed by the performance framework of routine information systems, it was primarily based on system theory. After conducting the analysis in Python and SPSS, the data was presented using descriptive statistics, such as means and standard deviations, and inferential statistics including regression analysis. The study observed that information timelines significantly moderated the connection between the technical factor and the integration of health management information systems.**

Keywords—Health management information system; cronbach alpha; moderator; information timelines

I. Introduction

Various forms of innovations have emerged due to the heightened international emphasis on reinforcing health systems to guarantee the delivery of higher-quality healthcare services. However, the objective has not been completely achieved. The [WHO], 2007 defines health system strengthening as a collection of programs and strategies that enhance one or more core elements of the health system, resulting in improved health. Each element within the World Health Organization framework holds significance. Conversely, rapid, precise, and relevant information is vital for effectively bolstering the health system to enhance its performance. It can also facilitate the sharing of common data and practices and enable the production and access of information in a real-time environment, as indicated by [1].

The objective is to furnish decision-makers with up-to-date information. Thus far, large-scale organizations, particularly in the manufacturing industry, have incorporated integrated ERP systems. These systems have been employed to streamline various business operations, encompassing sales, finance, production, and dispatching. It also facilitates the segmentation of healthcare functions in terms of information exchange and flow. A well-executed integrated information system offers numerous potential benefits and might even be indispensable for the survival and efficiency of an organization. The integration of IHMIS can assist healthcare organizations in managing their operations more efficiently, offering advantages such as inventory reduction, improved coordination in the supply chain, streamlined process flow, enhanced data analysis, decision-making using quality data, and improved patient care services [2]. Over the next decade, the ERP market is anticipated to be one of the application software industry's most rapidly expanding and crucial segments. Currently, it stands as one of the swiftest-growing software markets.

This remains true even though about sixty percent of information system implementation initiatives fail globally and across all organizations [3]. The author in [4] emphasizes that managers should take a proactive approach and actively encourage other users and staff members to adopt the system. Successful adoption would likely lead to improved patient care coordination and an elevated standard of care across all areas [5].

The health management information system (HMIS) initiative in India seeks to encourage healthcare professionals across various government levels about the potential of high-quality data to enhance patient care, representing an effort to leverage technology for public health improvement. Similarly, countries like Iran, Malawi, Kenya, and Uganda have adopted web-based systems to facilitate data access for decision-making. According to research, the management information provided by DHIS2, an open-source program that can be customized to include HIM functions, is found to be ineffective [6].

The methodology employed in this study is highly valuable. In this approach, the author utilized Independent Variables (IV), a Moderator, and a Dependent Variable (DV).

The structure of the remaining article is outlined as follows: Section II is Literature Review, Section III is Hypotheses, Section IV is Descriptive Statistics, Section V is Methodology, Section VI is Results and Discussion, and Section VII is Conclusion.

II. Literature Review

The idea of integrating information systems facilitates more efficient coordination and control of organizational activities and healthcare delivery was emphasized by Finnish researcher [7]. However, [7] does not delve into the facilitation of the integration process. A study conducted in African countries Ghana and Tanzania identified obstacles to HMIS integration, including limited capacity for data analysis and decision-making, redundant and parallel reporting, and communication channels. Establishing amicable connections and partnerships is crucial for the effective implementation of

IHMIS, as it necessitates the cooperation and commitment of numerous stakeholders. Previous studies have highlighted the benefits of information-sharing within healthcare partnerships [8].

A study conducted in China highlighted seven primary factors crucial to the success of the HMIS. These factors encompass strong motivation, a shared vision, a multidisciplinary implementation team, seamless integration with internal information systems, advanced legacy information systems and infrastructure, and adherence to industry standards [9]. To record and track population health data, most low-income nations still mostly rely on paper-based systems [10]. In regions with little infrastructure and resources, paper-based systems are frequently the most economical and useful modern option accessible. Coordinating data collection among government agencies can be challenging, especially when dealing with local healthcare service providers that perform similar tasks. However, donor policies typically encourage the implementation of vertical programs that support their own information systems and administrative structures [6].

Embracing modern technology represents just one of the various approaches healthcare organizations can utilize to enhance productivity and reduce costs. This investigation examines the technical aspect in line with the conceptual framework, considering:

1) IT infrastructure, comprises two distinct yet interrelated components: Technological infrastructure and human infrastructure. The technical infrastructure supports business applications (hardware, software, and data) with a collective set of tangible IT resources. b) Organizational and human skills, knowledge, expertise, commitments, values, and norms are all included in the human infrastructure. It is crucial to evaluate the sufficiency and accessibility of the human and technology infrastructure. It tackles the essential question of whether the information system would technically work, as mentioned by [11]. c) System interoperability. The primary emphasis of the laissez-faire approach is on organizational performance. Under laissez-faire leadership, employees are granted the autonomy to operate as they deem appropriate with minimal intervention. Leaders in this model assign responsibilities and make decisions, staying informed about corporate affairs and available for consultation as needed. However, they adopt a hands-off approach, enabling team members to work autonomously while meeting established organizational goals [12].

Data quality is delineated by four attributes: accuracy, timeliness, completeness, and consistency. Consistency refers to the level of correspondence between patient data recorded on patient cards and the registry. Meeting the reporting deadlines is the criteria used to gauge timeliness, as highlighted by [13]. The assessment of data quality is currently widely embraced in standard public health practices to ensure the reliability of data in Health Information Systems (HIS). To make evidence-based decisions, there is a need for both the demand and utilization of information. During their working hours, healthcare professionals collect a great deal of patient data; However, at the time of gathering, this data is hardly evaluated or utilized [14]. Therefore, an effective Health Management Information System (HMIS) should ensure that data collection is closely aligned with user requirements, encompassing only relevant data, and should possess processing capabilities that enable the easy retrieval of information, requiring only the essential data for expeditious analysis [13].

The results showed that Brazil does not use information effectively enough for planning or decision-making [13]. A prevalent deficiency in effectively using data to monitor service usage trends over time, thereby evaluating the impacts of policy and service delivery changes, is a significant vulnerability that impacts the entirety of the Sub-Saharan Africa [15].

An effective HMIS streamlines reporting by advocating for unified reporting to development partners and discouraging the implementation of parallel reporting systems whenever possible (WHO & ROWP, 2004). According to [16], work was classified into two main categories: (i) the type, capacity, specialization, and skills required for a specific task; and (ii) the diversity of knowledge and expertise from different individuals. This distribution of tasks highlights the potential to optimize the diverse talents and expertise of various healthcare workers, fostering the development of specialization. It also minimizes the time lost in duplicating tasks that could be efficiently handled by individuals based on their specific skill sets.

An organization's capacity to embrace technology was evaluated based on its utilization of Information and Communication Technology (ICT) and the promptness of information delivery, denoting the immediate and simultaneous exchange of information among all users. ICT facilitates electronic communication, information processing, and transmission, fostering the sharing of knowledge. This encompasses all forms of digital and analog information and communication technology, excluding non-electronic technologies. Examples of ICT in this context include computers, radios, televisions, phones, digital texts, and audio-video recording. The absorption and dissemination of innovation have served as the central analytical framework in investigating the adoption of information technology (IT) and information systems (IS). In their report, [17] underscores the significant role of an organization's technological proficiency and resources in its ability to adapt to new technologies.

According to study [18], healthcare professionals encounter challenges in implementing IT systems due to their complexity, resulting in a reliance on manual paper filing systems that compromise and mismanage information. According to study [19], information technology use and application are new ideas in modern developing-nation organizations. In situations involving emerging diseases and urgent health crises, where timely notification, investigation, and response can prevent widespread outbreaks and even global pandemics, as well as save lives, the need for precise information is particularly crucial [20]. The Government of [21], states that the main goals of the IHMIS are to improve clinical outcomes, decrease redundancy, boost efficiency, improve access and coordination, and strengthen connections between various care tiers and support services. Moreover, integrated healthcare harnesses the diverse skills and expertise of various healthcare professionals, along with crucial support services such as information management. Many people consider healthcare to be one of the foremost concerns for humanity, encompassing both societal and personal life objectives.

The correlation between information technology and healthcare is symbiotic. Accurate information is indispensable for the healthcare sector, and achieving precision in information is contingent on a person’s mental, physical, and psychological well-being [22]. The study in [23] objective is to aid both administrators and medical managers by elucidating the influence of the health information system on the decision-making process and highlighting the system's importance in facilitating these decisions.

III. Hypothesis

A hypothesis must be put to systematic testing or observation to see if it is true or not (Bradford, 2015). Only then can it be deemed scientific. The following hypotheses were examined in the study:

(H1): Behavioral factors play a substantial role in the integration of HMIS within the healthcare sector.

(H2): The leadership style of the health system is crucial to the successful integration of HMIS in healthcare facilities.

(H3): Healthcare organization’s HMIS integration is significantly impacted by technical aspects.

(H4): The timely delivery of information has a significant impact on the interaction between HMIS integration in healthcare organizations.

IV. Descriptive Statistics

An overview of the study's conclusions and descriptive statistics based on the data gathered are given in this section. There were 219 questionnaires distributed in total; 214 were collected, and five were discarded because the responses were not complete after the 214 surveys' data were processed, analyzed, and thoroughly examined for any missing values or anomalies.

According to the data presented in Table I, approximately 47% of the samples fell within the age range of 25 to 35, while 35% were aged between 35 and 45, with the remaining 18% being over the age of 45.

Table II indicates that 51% of the samples are male, while 49% are female.

Table III displays the sample's perspectives on the questionnaire statements regarding the data collection strategy. The average data collection strategy in the table is 3.38, indicating that the assessment of the data collection strategy was at a moderate level. The findings indicate that the statement means ranged from 3.28 to 3.49.

Table IV presents the sample's viewpoints on the questionnaire statements regarding the influence of the health system's leadership style on IHMIS. The average score for the impact of the health system's leadership style on IHMIS is 3.432, indicating a moderate level of influence. According to these findings, the statement means ranged from 3.15 to 3.74.

On the other hand, Table V displays the results of the Means & Sd. analysis for the statements concerning the Integrated HMIS Model.

Regarding the Integrated HMIS Model (Dependent variable), the statements were assessed using Means & Sd. The results are displayed in Table V.

The data in Table V reflects the sample's perspectives on the questionnaire statements related to the Integrated HMIS Model. The average for the Integrated HMIS Model was recorded as (3.635), indicating a moderate level of estimation. The results indicate that the statement means ranged from (3.37 to 3.95). Concerning the Information Timeliness (Moderating variable), Means & Sd. were used to evaluate the statements. Table VI presents the findings.

TABLE I. The Age Distribution of Participants' Frequency

Age	Frequency	Percent
25-35	100	46.7
35-45	75	35.0
45 & Above	39	18.2
Total	214	100.0

TABLE II. The Distribution of Participants' Gender Frequencies

Gender	Frequency	Percent
Male	109	50.9
Female	105	49.1
Total	214	100.0

TABLE III. Means and STD. Deviation of Data Collection Strategy

S#		Mean	Std. Deviation
1	All caregivers are committed to recording all collected information, whether manually or electronically.	3.41	1.236
2	Data collection adheres to the guidelines outlined in the templates.	3.29	1.166
3	Essential data is collected at each service point using the provided templates.	3.28	1.197
4	To ensure effective service delivery, each patient undergoes a series of well-organized processes.	3.39	1.115
5	The available HMIS enables seamless information exchange within the facility.	3.46	1.116
6	Multiple data sources are accessible within the premises.	3.49	1.056
7	HMIS is perfectly aligned with organizational structure.	3.36	1.064
Total		3.38	1.14

TABLE IV. MEANS AND STD. DEVIATION OF EFFECT OF THE HEALTH SYSTEM'S LEADERSHIP STYLE ON IHMIS

S#		Mean	Std. Deviation
1	Consistently provide basic outpatient care.	3.27	1.218
2	Deliver top-notch surgical services.	3.26	1.185
3	Focus is solely on promotive/preventive care.	3.26	1.056
4	Ensure satisfactory patient services.	3.3	1.156
5	Curative services meet the needs adequately.	3.48	1.091
6	Hospital operations are guided by established plans and objectives.	3.17	1.093
7	Consistently achieve targets within the designated time frame.	3.15	1.296
8	The available HMIS facilitates seamless information sharing with national referral hospitals.	3.56	1.004
9	Regularly exchange medical records with neighboring institutions throughout the county.	3.59	0.992
10	Hospital management team maintains regular communication with staff members.	3.71	1.027
11	Hospital's established structures enable easy access to healthcare for patients.	3.74	1.001
12	Each department within hospital has an adequate number of staff members.	3.56	1.002
13	Knowledgeable and proficient individuals are staffed at every service point within hospital.	3.56	1.076
Total		3.432	1.092

TABLE V. MEANS AND STD. DEVIATION INTEGRATED HMIS MODEL

S#		Mean	Std. Deviation
1	The data collected is consistently comprehensive.	3.75	1.053
2	Regular data audits ensure data quality.	3.65	1.062
3	Data collection, analysis, and utilization are integrated across all departments in the facility.	3.74	1.046
4	At facility, there is a significant demand for information to aid in decision-making processes.	3.73	0.988
5	The management frequently seeks evidence to validate the accuracy of reports used for decision-making.	3.61	1.099
6	Timely identification and correction of any deviations from planned activities is our norm.	3.64	1.01
7	Management team has implemented control mechanisms to ensure optimal organizational performance.	3.51	1.074
8	Willingly share information to aid disease prevention and control efforts.	3.38	1.012
9	Effective information-sharing has significantly improved cost savings within facility.	3.37	1.002
10	Relevant feedback for corrective action is promptly disseminated.	3.38	0.98
11	Management team at the facility often engages in benchmarking activities.	3.52	0.948
12	Consistently prepare reports meticulously and ensure they are well-organized.	3.4	0.962
13	Reports in facility meet the standards provided by the Ministry of Health.	3.43	0.926
14	HMIS-generated reports are instrumental in effecting changes within facility.	3.41	1.056
15	Via the different capabilities of HMIS, we receive abundant information from sub-county hospitals.	3.41	1.109
16	They remain informed about the local health situation and needs through frequent meetings with the county health department.	3.78	1.116
17	Maintain close cooperation with hospitals in sub-counties and health centers.	3.84	1.054
18	All sections, departments, and divisions work collaboratively to achieve organizational goals.	3.92	1.017
19	Submit regular reports to the sub-county Ministry of Health every week.	3.95	1.008
20	Facility utilizes one of the top HMIS programs available in the market.	3.93	1.027
21	The system conducts regular data backups.	3.88	0.959
22	Eectronic medical record management solution simplifies the process of record-keeping.	3.74	1.024
23	The data collected is consistently comprehensive.	3.64	0.986
Total		3.635	1.023

TABLE VI. MEANS AND STD. DEVIATION OF INFORMATION TIMELINESS

S#		Mean	Std. Deviation
1	Patients can easily access necessary information.	4.41	1.096
2	HMIS enables rapid reporting to DMIS.	4.31	1.053
3	Facility has established a dependable system for collecting and disseminating data.	4.38	1.118
4	HMIS primarily aids clinical healthcare workers in timely task completion.	4.54	1.051
Total		4.41	1.08

The findings in Table VI illustrate the perspectives of the respondents on the questionnaire statements related to Information Timeliness. The data demonstrates that the means of the statements ranged between 4.31 and 4.54, with Information Timeliness averaging at 4.41, signifying a high estimation level for Decision Making.

V. METHODOLOGY

Given that it aligns with the research objectives, the exploratory factor analysis approach is chosen for this study. Two main factors go into this decision. First of all, it seeks to reduce a large number of variables—132 in total—into a more manageable group of elements. Second, it follows the recommendations made by [24] in an attempt to preserve as much of the original variance as feasible.

Additionally, a more theoretical reason is considered when favoring EFA in the decision-making process. This reason relates to the correlation structure among measured variables, which is distinct from the goal of data reduction. It's worth noting that exploratory factor analysis is considered a data-driven approach because the researcher lacks a predetermined notion of the number of factors [25], It provides techniques for determining the appropriate number of factors and the configuration of factor loadings to be employed in subsequent hypothesis testing. When conducting an EFA, three fundamental steps are followed: assessing item correlations, extracting components, and rotating factors [26]. These aspects are elaborated upon in the following sections:

A. Rotated Factor Extraction

Fig. 1 illustrates that the number of elements obtained through rotation techniques is equivalent to those derived from the principal components method. Table VII presents the Kaiser-Mayer-Olkin measure of sampling adequacy, yielding a value of 0.765, which surpasses the suggested cutoff point of 0.6 (Hair et al., 2010). Moreover, Bartlett's Test of Sphericity yielded a significant result with a p-value of .000. Patterns with loadings exceeding 0.3 were identified using an analytical approach.

TABLE VII. KMO AND BARTLETT'S TEST

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.765
Bartlett's Test of Sphericity	Approx. Chi-Square	29348.966
	df	9045
	Sig.	.000

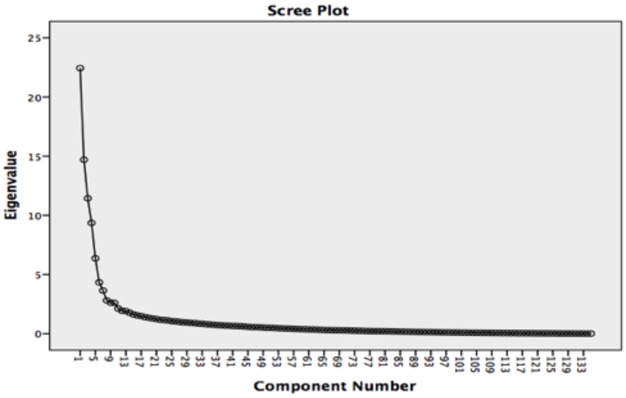


Fig. 1. Factors derived from the varimax and promax techniques with loadings exceeding 0.3.

B. Examining the Variables Associated with the Efficacy of Information Systems

Setting a loading cut-off threshold at 0.4 resulted in the successful extraction of nine components (see Fig. 2), with the essential output tables presented below. Notably, an impressive cumulative variance explanation of 83% is achieved, as indicated in Table VIII, and the KMO measure of sampling adequacy (see Table VII) is notably favorable. In conclusion, Table VIII pattern compilation distinctly identified nine components that underwent further in-depth examination.

TABLE VIII. KMO AND BARTLETT'S TEST

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.832
Bartlett's Test of Sphericity	Approx. Chi-Square	10740.155
	df	1830
	Sig.	.000

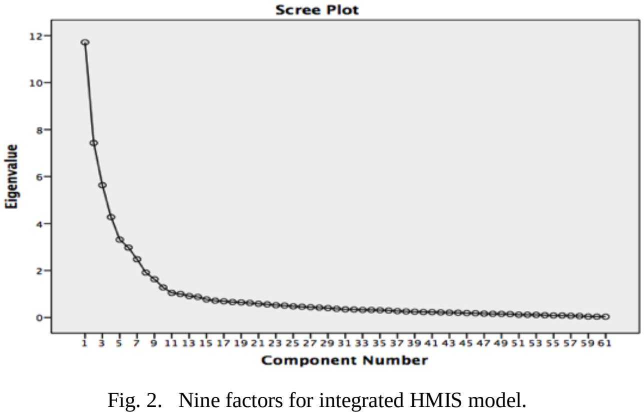


Fig. 2. Nine factors for integrated HMIS model.

C. Discussion on the Solution - Assigning Names to the Factors of the Independent Construct

The factor analysis results showed the extraction of nine factors that could be related to the independent construct's five dimensions. Four of the items in the pattern matrix had loadings that were less than 0.4, indicating that the original 132 variables were condensed to 51. Here is a list of these particular items.

TABLE IX. FACTORS AND VARIABLES ASSOCIATED WITH THE INTEGRATED HMIS MODEL

Dimensions	Factors 9	Variables 51
Information Quality	5	5
Effect of the Health System's Leadership Style on IHMIS	2	7
	5	6
Integrated HMIS Model	10	3
	1	11
	6	5
	7	4
Information Timeliness	9	4
Technical factor affecting IHMIS	3	9

Table IX provides a summary of the number of variables and factors retained for use in the study. Each extracted element underwent an individual evaluation to determine its characteristics and suitability for inclusion in the subsequent study. The pattern matrix revealed the distinctiveness of the factors, facilitating the identification of the weaker ones: Factor 8 encompassed five variables; Factors 2 and 5 comprised 7 and 6 variables, respectively. Factors 10, 1, 6, and 7 contained 3, 11, 5, and 4 variables, respectively. Lastly, Factors 3 and 9 included 4 and 9 variables, respectively.

D. *Reliability Analysis*

Data was collected and compiled using SPSS for analysis purposes. To ascertain the internal consistency reliability of the Health Management Information System (HMIS) integration, Cronbach Alphas were calculated. According to Nunnally (1978), an acceptable Cronbach Alpha level of at least 0.70 was defined for internal consistency reliability. Throughout the evaluation process, it was identified that several elements needed to be removed to enhance the instrument's reliability. Table X describes the findings of the factor's reliability.

TABLE X. FINAL FACTORS RELIABILITY FINDINGS

Constructs	Factor Name	Cronbach Alpha
Information Quality	Data Collection Strategy	0.906
Effect of the Health System's Leadership Style on IHMIS	Transformational Leadership	0.787
	Laissez-Faire Leadership	0.877
Integrated Hmis Model	Data And Information Quality	0.741
	Information Use	0.923
	Teamwork	0.879
	Technology Adoption	0.853
Information Timeliness	Information Timeliness	0.902
Technical Factor Affecting IHMIS	Technical Factor	0.732

E. *The Research Model*

Commonly referred to as a conceptual model, a conceptual framework is a visual representation that assists researchers in illustrating the anticipated cause-and-effect relationships [27]. The dimensions comprise Information Quality, Transformational Leadership, Laissez-Faire Leadership, and System interoperability as independent variables, Information timelines as a moderator, while Integrated HIMS serves as the dependent variable [28] and [29]. Using a multiple regression model, this study sought to determine the impact of the three independent variables (P_1, P_2, and P_3) and the single moderating variable (Q) on the dependent variable (R) through HMIS integration efforts. The study's objectives were met through the hierarchical examination of variables, following the procedures outlined in the Fig. 3 conceptual framework proposed by [30].

$$R = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \beta_3 P_3 + \beta_4 Q + \beta_5 P_1 Q + \beta_6 P_2 Q + \beta_7 P_3 Q$$

Let Y represent the HMIS integration, where β0 stands for the intercept, and β_1, β_2, and β_3 represent the slope coefficients, indicating the relationship between the respective independent variables and the dependent variable. P_1 denotes the behavioral component of health professionals, P_2 represents the influence of the health system's leadership style on IHMIS, and P_3 corresponds to the technical factor. Each hypothesis was carefully tested and evaluated to ensure its alignment with the study's objectives.

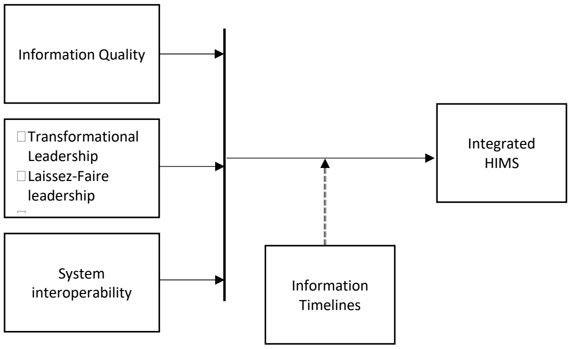


Fig. 3. Conceptual framework.

The model examined the influence of the independent variables (P_1: healthcare professionals' behavior, P_2: the influence of the health system's leadership style on IHMIS, and P_3: the technical aspect) on the dependent variable (R: Integrated HMIS Model). Before considering any moderating effects of information timeliness, the model included the standard predictors of integration in healthcare organizations.

VI. RESULTS AND DISCUSSION

The multiple regression model presented in Table XI indicated that Information Quality (β2 = -0.1053, P > .005) and Technical factor (β4 = 0.0834, P > .005) are not significant contributors to the integration of HMIS in a combined relationship. However, Health System's Leadership Style (β3 = 0.4654, P < .005) emerged as a significant influence on HMIS integration in a combined relationship. This implies that the crucial factor predicting HMIS integration is leadership style, with other factors being less impactful.

The presented results outline the outcome of an Ordinary Least Squares regression analysis in Table XII. The purpose of the model is to predict the values of the dependent variable by considering the independent variables (IV1, IV2, IV3) and their interactions with the moderator (IV1_M_interaction, IV2_M_interaction, IV3_M_interaction). Assuming that all other variables stay constant, the coefficients for each variable show how the dependent variable is expected to vary for every unit change in the associated independent variable.

The model explains 69.9% of the variance in the dependent variable, indicated by an R-squared of 0.699.

The model demonstrates overall statistical significance, as evident from the low p-value of 2.36e-50 and the high F-statistic of 68.48. IV1 does not appear to be statistically significant, indicated by its high p-value of 0.614 and a coefficient of -0.1053. IV2 is deemed statistically significant due to its coefficient of 0.4654 and a relatively low p-value of 0.005. IV3 does not exhibit statistical significance, as indicated by its high p-value of 0.572 and a coefficient of 0.0834.

The interaction terms IV1_M_interaction, IV2_M_interaction, and IV3_M_interaction are denoted by coefficients of 0.1181, -0.1320, and -0.0395, respectively. IV1_M_interaction, IV2_M_interaction, and IV3_M_ interaction were found to be statistically significant in the context of the provided data, however, IV3_M_interaction was not, based on their respective p-values. These findings imply that the moderator variable has a significant impact on how the independent variables affect the dependent variable. Furthermore, the research shows that IV2 has a significant and unique effect on the dependent variable, in contrast to IV1 or IV3.

DV=IV1_M_interaction - 0.1320 * IV2_M_interaction - 0.0395 * IV3_M_interaction + 0.1181 * Moderator + 0.0804 - 0.1053 * IV1 + 0.4654 * IV2 + 0.0834 * IV3 + 0.7164 * Moderator.

The link between the dependent variable and multiple independent variables (IV1, IV2, and IV3), as well as their corresponding coefficients, are shown in the regression equation. The influence of the Moderator variable on the relationship between the independent and dependent variables is also accounted for in the equation.

TABLE XI. SUMMARY OF THE TESTED HYPOTHESES' OUTCOMES

No.	Variable	p-value	Direction	Deduction
1	Information Quality	0.614	Negative	Do not reject Null
2	Effect of the Health System’s Leadership Style on IHMIS	0.005	Positive	Reject Null
3	Technical Factors affecting IHMIS	0.572	Positive	Do not reject Null
4	Information Timeliness	0.000	Positive	Reject Null
5	Information Quality* Information Timeliness	0.026	Positive	Reject Null
6	Effect of the Health System’s Leadership Style on IHMIS * Information Timeliness	0.004	Negative	Reject Null
7	Technical factor affecting IHMIS* Information Timeliness	0.327	Negative	Do not reject Null

TABLE XII. ORDINARY LEAST SQUARES REGRESSION ANALYSIS

	coef	std err	t	P>ltl	[0.025	0. 9751]
const	0.6804	0.643	0.125	0.901	-1.188	1.348
IVI	-0.1053	0.208	-0.506	0.614	-0.516	0.305
IV2	0.4654	0.162	2.872	0.005	0.146	0.785
IV3	0.0834	0.147	0.566	0.572	-0.207	0.374
Moderator	0.7146	0.190	3.752	0.000	0.339	1.09
IVI_M_interaction	0.1181	0.052	2.25	0.026	0.015	0.222
IV2_M_ interacion	-o.1320	0.046	-2.874	0.004	-0.223	-0.041
IV3_M_interaction	-0.0395	0.040	-0.982	0.327	-0.119	0.04

The anticipated change in the dependent variable for a one-unit increase in the associated independent variable, assuming other variables remain the same, is displayed by the coefficients for each independent variable (IV1, IV2, IV3). Moreover, IV1_M_interaction, IV2_M_interaction, and IV3_M_interaction are interaction terms that show how the moderator changes the associations between the independent and dependent variables. Positive coefficients point to an amplification impact, whilst negative coefficients point to a damping effect.

The investigation unveiled a strong and positive connection between the timeliness of information and the integration of HMIS. Furthermore, it was observed that a powerful leadership style notably improved the functionality of health management information systems, leading to enhanced patient outcomes.

VII. Conclusion

In terms of the organizational dimension, it was imperative to enhance, endorse, and adjust data collection methods to meet the dynamic requirements of the healthcare system. At that time, the healthcare system used a lot of different data collection tools, which led to laborious work and weariness among data and information workers. The integration of HMIS was notably associated with the Health System's Leadership Style, suggesting that healthcare organizations could achieve operational efficiency by refining their health system leadership style to align with environmental shifts. The results of the study showed a strong and positive association between information timeliness and HMIS integration. Overall, the study's model explained 69.9% of the variability in HMIS integration, demonstrating its efficacy in answering important questions about the critical elements of HMIS integration in healthcare facilities. It was observed that a strong leadership style greatly enhanced the functionality of health management information systems, which in turn improved patient outcomes.

In future work, the current study can also be mapped with two important independent variables named as Information Sharing and Complexity of Project Management.

References

[1] Wickramasinghe V, Karunasekara M. Perceptual differences of enterprise resource planning systems between management and operational end-users. Behaviour & Information Technology, 31(9), 873-87, 2012.

[2] Seth M, Goyal DP, Kiran R. Development of a model for successful implementation of supply chain management information system in Indian automotive industry. Vision, 19(3), 248-62, 2015.

[3] Maas JB, van Fenema PC, Soeters J. ERP system usage: The role of control and empowerment. New Technology, Work and Employment, 29(1), 88-103, 2014.

[4] Yen HR, Hu PJ-H, Hsu SH-Y, Li EY. A multilevel approach to examine employees' loyal use of ERP systems in organizations. Journal of Management Information Systems, 32(4), 144-78, 2015.

[5] Hwang W, Chang J, LaClair M, Paz H. Effects of integrated delivery system on cost and quality. Am J Manag Care, 19(5), e175-e84, 2013.

[6] Kiberu VM, Matovu JKB, Makumbi F, Kyozira C, Mukooyo E, Wanyenze RK. Strengthening district-based health reporting through the district health management information software system: the Ugandan experience. BMC medical informatics and decision making, 14(1), 1-9, 2014.

[7] Koskinen KU. Knowledge integration in systems integrator type project - based companies: a systemic view. International Journal of Managing Projects in Business, 5(2), 285-99, 2012.

[8] Evans MF, Thomas S. Implementation of an integrated information management system at the National Library of Wales: A case study. Program: electronic library & information systems, 41(4), 325-37, 2007.

[9] Lu X-H, Huang L-H, Heng MSH. Critical success factors of inter-organizational information systems—A case study of Cisco and Xiao Tong in China. Information & management, 43(3), 395-408, 2006.

[10] Mbondji PE, Kebede D, Soumbey-Alley EW, Zielinski C, Kouvividila W, Lusamba-Dikassa P-S. Health information systems in Africa: descriptive analysis of data sources, information products and health statistics. Journal of the Royal Society of Medicine, 107(1_suppl), 34-45, 2014.

[11] Odhiambo-Otieno GW. Evaluation of existing district health management information systems: a case study of the district health systems in Kenya. International journal of medical informatics, 74(9), 733-44, 2005.

[12] Humaidi N, Balakrishnan V. Leadership styles and information security compliance behavior: The mediator effect of information security awareness. International Journal of Information and Education Technology, 5(4), 311, 2015.

[13] Teklegiorgis K, Tadesse K, Terefe W, Mirutse G. Level of data quality from Health Management Information Systems in a resources limited setting and its associated factors, eastern Ethiopia. South African Journal of Information Management, 18(1), 1-8, 2016.

[14] Gillingham P, Graham T. Designing electronic information systems for the future: Social workers and the challenge of New Public Management. Critical Social Policy, 36(2), 187-204, 2016.

[15] Nyamtema AS. Bridging the gaps in the Health Management Information System in the context of a changing health sector. BMC medical informatics and decision making, 10, 1-6, 2010.

[16] Gulick L, Urwick L. Notes on the Theory of Organization: Columbia University. Institute of Public Administration; 1937.

[17] Shiels H, McIvor R, O'Reilly D. Understanding the implications of ICT adoption: insights from SMEs. Logistics information management, 16(5), 312-26, 2003.

[18] Boone D, Cloutier S, Lins S, Makulec A. Botswana's integration data quality assurance into standards operating procedures: adaptation of the routine data quality assessment tool. Chapel Hill: Measure Evaluation, 2013.

[19] Shiferaw AM, Zegeye DT, Assefa S, Yenit MK. Routine health information system utilization and factors associated thereof among health workers at government health institutions in East Gojjam Zone, Northwest Ethiopia. BMC medical informatics and decision making, 17(1), 1-9, 2017.

[20] Friberg IK, Kinney MV, Lawn JE, Kerber KJ, Odubanjo MO. Sub-Saharan Africas Mothers. Newborns, and Children: How, 2010.

[21] Canada Go. An Overview of Progress and Potential in Health System Integration in Canada. Health Canada, 2002, 2002.

[22] Al-Adwan MM. The Impact of Information Technology on Health Sector Performance: Case Study–Saudi Arabia Health Sector. The Pacific Journal of Science and Technology, 2016.

[23] Alfawaz K, Alharthi S. The Role of MIS in Enhancing the Decision-making Process in Hospitals and Health Care Sectors: Case Study of AL-HADA Military Hospital in AL Taif, KSA. Egyptian Computer Science Journal, 43(2), 2019.

[24] Conway JM, Huffcutt AI. A review and evaluation of exploratory factor analysis practices in organizational research. Organizational research methods, 6(2), 147-68, 2003.

[25] Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychological methods, 4(3), 272, 1999.

[26] Lamoureux EL, Pallant JF, Pesudovs K, Rees G, Hassell JB, Keeffe JE. The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and Rasch analysis. Investigative ophthalmology & visual science, 48(3), 1001-6, 2007.

[27] Van der Waldt G. Constructing conceptual frameworks in social science research. TD: The Journal for Transdisciplinary Research in Southern Africa, 16(1), 1-9, 2020.

[28] Kyalo CK. Integration of health management information systems in health care organizations in Kenya 2018.

[29] Maritim TK. Project Management Information Systems and Decision-making in a Multi-project Environment (2022).

[30] MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychological methods, 7(1), 83, 2002.

A Sophisticated Deep Learning Framework of Advanced Techniques to Detect Malicious Users in Online Social Networks

Sailaja Terumalasetti[1], Reeja S R[2]
School of Computer Science and Engineering, VIT-AP University, Amaravati, India

***Abstract*—Malicious user detection is a cybersecurity exploration domain because of the emergent jeopardies of data breaches and cyberattacks. Malicious users have the potential to detriment the system by engaging in unauthorized actions or thieving sensitive data. This paper proposes the dual-powered CLM technique (Convolution neural networks and LSTM) and optimization technique, a sophisticated methodology for distinguishing malicious user behavior that assimilates LSTM and CNN, and finally optimization technique to enhance the results. A genetic algorithm is used to augment the model's capability to perceive altering and nuanced malicious performance by fine-tuning its parameters. Due to the rising vulnerabilities of data breaches and cyber-attacks, malicious user identification in OSN (Online Social Networks) is a significant topic of research in cybersecurity. The proposed technique pursues to ascertain anomalous user behavior patterns by assessing vast quantities of data generated by digital systems with CLM and optimizing detection accuracy with genetic algorithms. On a public dataset of social media bot dataset, a twibot-20 dataset comprehending user activity data, was explored to measure the performance of the suggested methodology. The outcomes demonstrated that, in comparison to conventional machine learning algorithms like SVM and RF, which respectively obtained 92.3% and 88.9% accuracy, our technique, had a better accuracy of 98.7%. Moreover, the other metrics measures were assessed, and the proposed technique outperformed traditional machine learning algorithms in each situation.**

***Keywords*—Online social networks; malicious user behavior; convolution neural networks; long short-term memory; genetic algorithm**

ABBREVIATIONS

Acronyms	Definition
OSN	Online Social Networks
CNN	Convolution Neural Network
LSTM	Long Short-Term Memory
GA	Genetic Algorithm
CLM	Convolutional Neural Network and LSTM
NLP	Natural Language Processing
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
Acc	Accuracy
Prec	Precision
Rc	Recall
$F1_s$	F1- Score

I. INTRODUCTION

Online social networks have turned out to be an indispensable element of our everyday life. Platforms like Facebook, Twitter, Instagram, and LinkedIn have transformed the way we intermingle, altercation data, and associate with people.

For cybersecurity professionals, ascertaining malicious users in online social networks (OSN) presents a perplexing task. People might now enthusiastically interact with friends and families, discuss their views and opinions, and even conduct business online, acknowledging the upsurge of social media. Online social networks (OSNs) have turned out to be a crucial part of the contemporary era, endorsing connectivity and information sharing. However, as these platforms are exposed, they are probable to diverse sorts of misuse, including malevolent user activity. The research deals with the crucial theme of detecting and mitigating malicious user behavior. Online social networks (OSNs) are virtual communities that allow individuals to associate and communicate with one another on a certain topic or just "hang out"[1].

With billions of handlers worldwide, OSNs have turned out to be an indispensable component of modern civilization. Individuals are progressively using OSN sites due to the rapid growth of Web 2.0 technology. The rise of malevolent individuals in online social networks, on the other hand, has become a substantial concern for both users and researchers. Criminal hackers recurrently exploit social media to spread spam and malware, which is acknowledged as social malware. These destructive users will not "fit" into any of these classifications because they have mutual friends and interests and develop gigantic communal networks. The advent of detrimental handlers in online social networks is an intensifying basis of concern for users. According to one assessment, the number of fraudulent social media profiles generated grew by 100% in the first half of 2020. According to another survey, the amount of social media phishing attacks grew by 500% in the first quarter of 2021. These statistics lay emphasis on the prominence of detecting and preventing fraudulent users in online social networks.

Security is of utmost consequence in the contemporaneous era, since the majority of our private and sensitive data is stockpiled digitally. Malicious users have the potential to harm the system by engaging in unauthorized actions or stealing sensitive information [3]. Access restrictions, intrusion detection systems, and firewalls are instances of traditional security measures that can assist in preventing attacks to some extent but are not precisely operational when it comes to malicious user detection. Algorithms in machine learning have been used to analyze user activity and detect anomalies. However, the accuracy of these algorithms is mostly determined by the prominence and volume of training data.

A malicious user utilizes a computer system or network intending to cause harm, steal data, or disrupt normal operations [1]. Malicious users may have numerous intentions, comprising of financial gain, retaliation, or political involvement. They may use a variety of strategies to accomplish their goals, encompassing malware, phishing, social engineering, and exploiting vulnerabilities in software and hardware.

Analyzing user behavior is one procedure for identifying malevolent users. It could be capable of flagging suspicious behavior and more research by discerning an eye on user activity patterns and perceiving abnormalities. CNN and LSTM networks are instances of machine learning techniques that possibly will be used to automatically analyze big datasets of user behavior and predicament patterns that can be suggestive of harmful conduct [2]. By looking for the ideal set of hyper parameters, genetic algorithms (GAs) may be employed to improve the enactment of the archetype.

Malicious users pose a severe threat to entities, governments, and organizations. They have the proficiency to steal private information, jeopardize the security of systems, and harm a company’s reputation and brand. Therefore, it is essential to have effective techniques for identifying and reducing the actions of harmful users [3]. The upsurge of these daily threats over the past ten years is the main cause for concern for data security. Fig. 1 illustrates the tendency of the threats in the past decade.

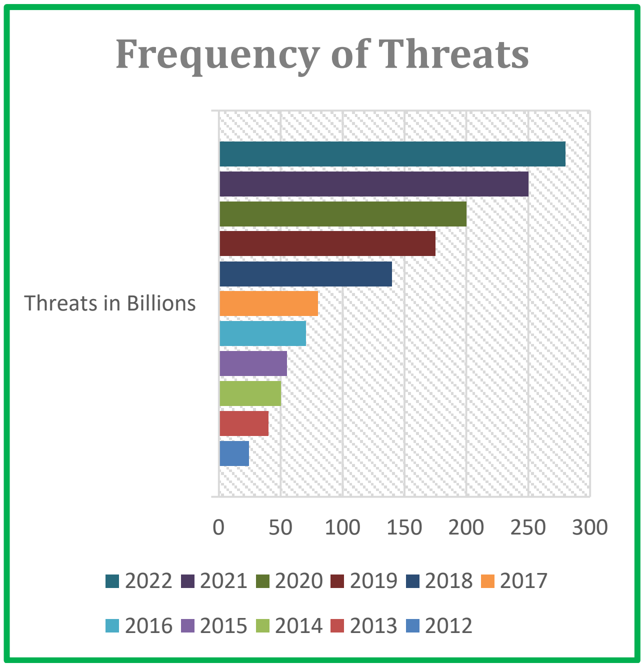


Fig. 1. Frequency threats in real-time.

A unique approach is envisioned with a dual-powered CLM (Convolution neural networks and LSTM) and optimization technique. The amalgamation of deep learning and evolutionary computation provides the technique with the adaptive competencies vital to safeguard OSNs. The suggested method is evaluated on a user activity dataset in OSN, and the outcomes are illustrious from those of conventional machine learning techniques [4].

The motivation for the proposed CLM and Optimization method distinguishes hazardous users to improve security and defend against cyberattacks. Exploiting system vulnerabilities, attainment unauthorized access, stealing sensitive data, and interrupting system operations can detriment people and companies. Firewalls and antivirus software don't always stop complex attacks, thus modern methods are obligatory to detect and preclude them [13].

A. *Organisation of the Paper*

The paper encompasses the subsequent subheadings: Section II - Literature Review, Section III – Proposed Methodology, Section IV - Experimental Evaluations and Results, Section V - Conclusion and References.

II. Literature Review

Deep learning neural networks of the variation known as CNNs are frequently engaged in processing images and videos. They have been revealed to be incredibly efficacious in resolving stimulating computer vision issues comprising segmentation, object identification, and picture categorization. The vital principle of CNNs is to extract information from pictures using convolutional filters and then to categorise or determine objects using these characteristics [5]. CNNs have revolutionised the field of computer vision and made it possible for a variety of applications, from self-driving cars to medical imaging. CNN has significantly augmented its popularity in voice and picture recognition tests. It captures spatial and temporal tendencies in data since it is built on the notion of native connectedness and shared weights. When creating a CNN model, data inputs like images or data categorizations are deployed through numerous of layers of convolution, pooling, and activation functions. Ensuing this, fully linked layers that dispense the response into numerous classifications acquire the yield of these layers.

CNN has significantly amplified its popularity in voice and picture recognition tests. It captures spatial and temporal tendencies in data since it is built on the notion of native connectedness and shared weights. When creating a CNN model, data inputs like images or data sequences are deployed through numerous layers of convolution, pooling, and activation functions. Ensuing this, fully linked layers that distribute the response into several classifications acquire the yield of these layers [7].

The LSTM variance of the recurrent neural network (RNN) properly resolves the vanishing gradient problem that concerns regular RNNs. [6]. The vanishing gradient problem occurs when gradients get tinier as they propagate over time, making training the network on lengthy sequences challenging.

This problem will be resolved by LSTM, which has a particular form of memory cell that can store information for longer. Three gates govern the cell: the input gate, the forget gate, and the output gate. The forget gate standardizes the retention of preceding data, the input gate controls the flow of new information into the cell, and the output gate regulates the cell's output.

In an extensive assortment of applications, including speech recognition, machine translation, and NLP (natural language processing), LSTM has been illustrated to be effective. It has also been used for anomaly detection and time-series prediction jobs, where it may discover temporal relationships and long-term trends in data [8].

A heuristic optimization method based on natural selection and evolution is referred to as the Genetic Algorithm (GA). It is used to address optimization issues that require determining the optimal parameter combination for a given objective function. The GA generates a population of candidate solutions, known as chromosomes. Each chromosome is composed of a series of genes that represent various parameters of the issue being optimized. These parameters can include any form of data, including numerical values, Boolean values, and texts. Subsequently, the GA evaluates the fitness value of the respective chromosome in the population using the objective function. The fitness value assesses how successfully the chromosome resolves the issue. The GA then chooses the population's top chromosomes to serve as the parents of the following generation.

Employing genetic operators like crossover, mutation, and selection to the parents, the next generation gets generated. As opposed to mutation, which involves altering certain genes in a chromosome at random, crossover involves transferring genes between two chromosomes to produce new progeny. In selection, the population's finest chromosomes are chosen to serve as the parents of the following generation [27] [29].

Using the conceptions of natural selection and evolution, GA is a persuasive optimization technique that may unearth the preeminent responses to thought-provoking issues. It is comprehensively utilized through several disciplines, including computer science, engineering, and finance.

Malicious user detection prominence is evolving in the contemporary era because of security theft and data privacy. The information in this digital world has to be secure. The identification of malicious users is an important part of cybersecurity [9] [30]. User behaviour analysis (UBA) is a technology that employs machine learning and data analytics to detect abnormal conduct that might suggest a malevolent user. In this literature review, we will look at some current studies on detecting illegitimate users using UBA.

A machine learning-based approach to identifying illicit behaviour based on host process data was proposed by Han et al. [2]. The authors analysed user behaviours and identified abnormal behaviour using big data. The study is shown that UBA can be an effective method for detecting harmful activities. A user behaviour analysis system that utilises data analytics and machine learning to detect and differentiate that exists between malicious and genuine users was introduced by Ranjan and Kumar [6]. The authors analysed user behavioural data using multiple machine-learning methods to identify unusual behaviours. The study demonstrated that UBA can be an expedient method for detecting malicious users. Tanuja et al. [12] proposed a machine learning technique for identifying fraudulent social network users. The authors analysed user activity data using multiple machine-learning methods to identify abnormal conduct that may advocate a deceitful user. To identify various anomalous user behaviours and lessen their negative impacts, statistical analysis was done. To find unusual conduct that may point to a malevolent user, the authors performed statistical analysis [10]. Several patents pertaining to the detection of malicious users are accessible on Google Patents, including a framework for mobile advanced persistent threat detection, a deep learning method for detecting covert channels in the domain name system, and a technique for detecting insider and masquerade attacks by identifying malicious user behaviour [11] [12].

III. Proposed Methodology

A. System Model

System model for malicious user detection through user behavior for CLM and optimization technique. The Fig. 2 gives an overview of the system. The data collection and preprocessing module, the CLM and optimization technique, and the evaluation module encompass the classification model for malicious user detection through user behavior for CLM and optimization technique.

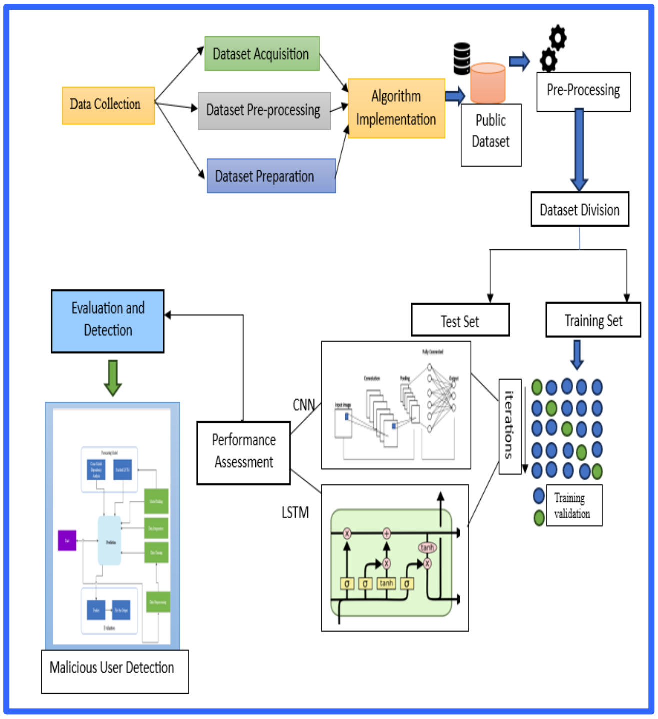


Fig. 2. Schematic diagram of proposed methodology.

The problem with devising malicious user detection employing CLM and optimization technique approach is to develop an artificial intelligence model that can precisely distinguish malicious users through user behavior information composed from a miscellaneous variety of sources. The model should be able to handle large datasets, noisy data, and a wide variety of malicious behavior types, including network attacks, system intrusions, and user impersonation. The objective is to develop a prototype that can be deployed in a real-world setting to detect and prevent malicious user behavior before it can cause damage to users or systems. The system aims to afford a reliable and precise methodology for identifying malevolent users by scrutinizing their behavioral patterns. The system attempts to capture both the spatial and temporal aspects of user behavior data by utilizing the capabilities of CNN and LSTM neural networks [17] [18]. In order to upsurge the model's performance and optimize its parameters, the Genetic Algorithm is also used.

A dataset of user behavior that comprises elements like login patterns, session length, transaction history, and other appropriate data is used as the system's input. The data is pre-processed by the method in order to normalize and encrypt it for neural networks. The CNN module of the classification pulls spatial characteristics from the input data, while the LSTM component captures the temporal relationships [26]. The CNN and LSTM model is then trained expanding the training dataset [19] [20]. The Genetic Algorithm is used to optimize the model, which scrutinizes various amalgamations of hyper parameters to classify the optimal collection of parameters that maximizes the detection accuracy [23].

1) Data collection: The data collection and preparation module is responsible for gathering user behavior data and converting it into a format that can be used by the AIMDS model. This module collects data from many sources, such as network traffic logs, user input logs, and system logs, and then pre-processes the data to eliminate noise, missing values, and other abnormalities [16].

a) Dataset acquisition: A large-scale dataset comprehending user behaviour information is attained from a reliable internet platform. The dataset encompasses a diversity of features such as user activities, timestamps, and session information.

b) Data Pre-processing: Pre-processing the dataset to eradicate excessive or redundant characteristics, manage missing values, and normalize the data [14]. Pre-processing processes may include feature selection, data purification, and categorical variable encoding.

c) Data preparation: The pre-processed data is consequently prepared for model training. The data has been fragmented into training, validation, and testing sets to accomplish this. The training set is utilized to train the prototypical, the validation set usage to fine-tune the hyper parameters, and the testing set is used to assess the aftermath of the model.

Preprocessing the input data entails filtering and normalizing the user behavior data to filter the noise and insignificant data as the first stage. The feature extraction layer utilizes the input data to extract useful characteristics that may be utilized for further exploration once the preprocessed data has been passed through it [21]. The Algorithm 1 provides the overview of the data preprocessing after the assemblage of the dataset has to endure a sequence of steps to further process.

Algorithm 1 : Data Preprocessing

```
Initialize
BEGIN
Step 1: Load the Dataset
Step 2: Handle the missing values
Replacing with Mean or Median values
Step 3: Normalize the features
Step 4: Splitting the dataset
Divide the Dataset
1.Training Dataset
2. Testing Dataset
Step 5: Feature Selection and Feature Extraction
Step 6: Handling the time series data
Step 7: Data augmentation
Step 8: Finalize the pre-processed dataset
End
```

2) Algorithm implementation: The CLM and optimization technique model is in possession of assessing the pre-processed user behavior data and determining whether or not a certain user is acting maliciously. This model is made up of two key parts: the CNN and LSTM layers, which extract features from user behavior data, and the genetic algorithm, which optimizes the CLM technique (Convolution neural networks and LSTM) model's parameters to enhance its accuracy [24][25].

Detecting malicious user behavior using the dual-powered CLM technique and an optimization technique approach involves several algorithms formulas and techniques.

a) Architecture: The CLM and optimization model architecture is intended based on the three algorithms. The CNN layer accumulates spatial characteristics from data, the LSTM layer captures the temporal dynamics of user behaviour [15], and the GA layer optimizes the model's hyper parameters.

b) Training: The CLM and optimization model is trained using the prepared data. The model is trained on the training set, then it is validated on the validation set. During the training phase, the loss function is minimized using optimization techniques such as stochastic gradient descent or Adam optimization.

c) Hyperparameter tuning: The hyper parameters of the model are optimized using the GA. The GA is used to explore the hyper parameter space for the optimum hyper parameter amalgamation that maximizes the model's performance. The GA's fitness function is based on evaluation measures such as Acc, Prec, Rc, and $f1_s$.

d) CLM Algorithm: The Algorithm 2 gives the details of initialization of the convolution layer parameters and applying activation function. The scientific formulation for the CNN component involves convolutions and pooling operations. Let's symbolize the input data as X, the convolutional layer output as C, and the pooling layer output as P. The Eq. (1) and Eq. (2) gives the desired outcome.

$$C = relu(conv(X)) \tag{1}$$

$$P = \max\_pool(C) \tag{2}$$

Algorithm 2: CLM

```
BEGIN
Initialize CNN parameters
f = filtersize
n=numoffilters
d= dropoutrate
fz= filtersizes
Define CNN
Inputlayer=input (shape= (input_shape))
Convlayers= []
For f in fz:
Convlayer= Conv1D (filters=n, kernelsize=activation='relu') (input_layer)
Poollayer    =MaxPooling1D(poolsize=x) (convlayer)
Convlayers.append (poollayer)
mergedlayer = Concatenate (axis=1) (convlayers)
flattenlayer   = Flatten () (mergedlayer)
dropoutlayer = Dropout (dropoutrate)(flatten_layer)
outputlayer=
Dense(numclasses,activation='softmax') (dropoutlayer)


Compile and train the model
END
```

e) Optimization algorithm: The Algorithm 3 describes the Genetic algorithm of initialization of population size, evaluation of fitness, probabilistic selection to evaluate the best solution. The fitness function in the genetic algorithm analyses the quality of each potential solution (chromosome). The fitness value is determined by the problem's purpose and can be a combination of metrics such as accuracy, precision, recall, or F1-score. The fitness function directs the genetic algorithm's selection, crossover, and mutation processes.

Algorithm 3: Genetic Algorithm

```
BEGIN
GA(Ft,Ft_th,s,f,m)
Ft: Fitness function assigns evaluation score
Ft_th: Fitness threshold
s: hypotheses to be included
F: fraction of population to be replaced
m: mutation error
Step 1: Initialization
Define population size
                $P \leftarrow$ Generate P hypothesis
Step 2: Evaluation
Compute fitness
Calculate fitness score
Step 3: Selection
The probability $Pr(s_i)$ is
                $Pr(s_i) = \frac{Fitness(s_i)}{\sum_{j=1}^{p} Fitness(s_i)}$
Step 4: Crossover
Select pair of hypothesis from P
For each pair produce offspring by applying crossover
Step 5: Mutation
Choose members with uniform probability
Step 6: Update
                $P \leftarrow P_s$
Step 7: Evaluate
Retrieve the best solution
END
```

3) Evaluation and detection: The evaluation module is in charge of establishing the CLM and optimization technique model is accurate and successful at detecting harmful user behavior. This module often consists of testing the model's performance on a test set of data and comparing its accuracy, precision, recall, and F1 score to other cutting-edge machine learning models like SVM and Random Forest.

a) Evaluation: The CLM and optimization technique efficacy is assessed using the testing set. Some of the assessment metrics used include Acc, Prec, Rc, and $f1_{s.}$ The results are compared to other cutting-edge methodologies to assess the efficacy of the suggested methodology.

b) Malicious user detection: Based on their conduct, the trained proposed model CLM and optimization technique are applied to detect malicious users. The model accepts data on user behaviour as input and produces the probability of the individual being malevolent. Based on the output prospect, a threshold is defined to identify people as malicious or non-malicious. The methodology's architecture is depicted in Fig. 3.

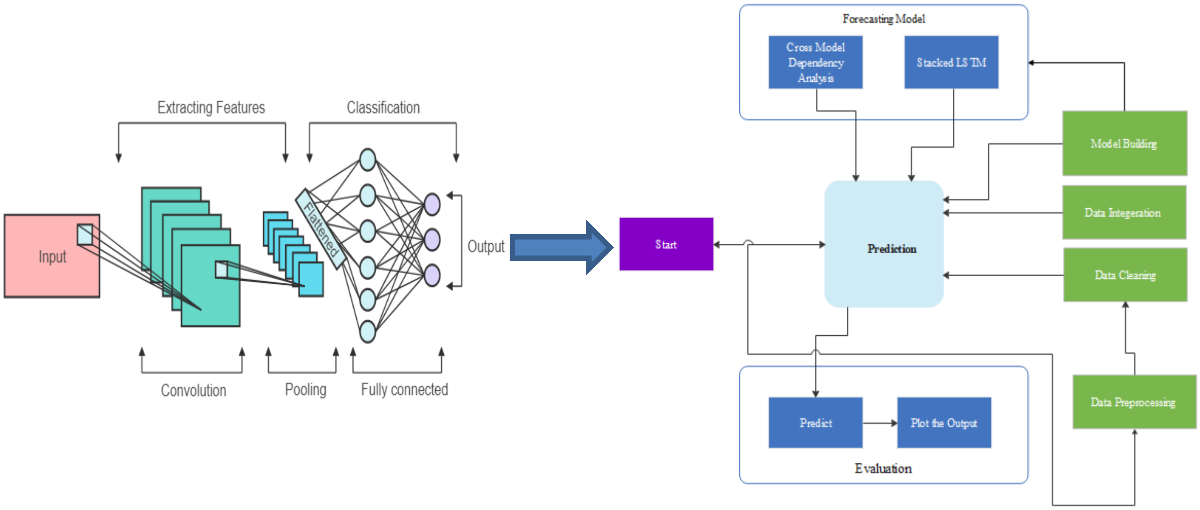


Fig. 3. Architecture to detect malicious user.

IV. Experimental Evaluations and Results

A. Dataset Description

The TwiBot-20 dataset, specifically designed for social media bots, serves as a substantial and all-encompassing standard for detecting Twitter bots. The purpose is to stimulate the difficulties posed by a small dataset size and accurately reflect both actual people and Twitter bots found in the real world. The collection comprises 229,573 people, 33,488,192 tweets, 8,723,736 user property pieces, and 455,958 follow relationships. It comprises a comprehensive range of automated accounts and authentic users to more accurately depict the Twitter community as it exists in reality. The dataset contains three different types of user information, which may be used for both classifying individual users into two categories and developing community-aware methods. The three modalities are semantic information, property information, and neighborhood information. The TwiBot-20 dataset is accessible for academic research objectives and is hosted by the Bot Repository [22]. This benchmark is one of the most extensive collections of Twitter bot detection data available. It obliges as an accommodating tool for training and assessing the proposed model that aim to identify harmful users in online social networks, specifically in the context of Twitter bot identification.

Considering the objective of achieving optimal performance in identifying harmful user activity, it is vital to conduct experiments and prudently tune the settings. The properties of the dataset, the kind of malicious activity, and the computational resources that are available for training and optimization all have a role in the selection of parameters. When trying to fine-tune these parameters in an efficient manner, it is frequently prerequisite to do iterative refinement based on performance data and domain expertise.

B. Experimental Results

Numerous indicators may be used to measure the success of a system built to identify harmful user behaviour using CLM and optimization techniques [28] [31]. Considering the frequently used assessment metrics.

C. Accuracy

Accuracy assesses the overall efficacy of the model's predictions. It computes the proportion of correctly identified cases (both harmful and non-malicious) in the dataset to the total number of occurrences. A higher level of accuracy suggests superior performance. Eq. (3) can be used to evaluate the accuracy. The Fig. 4 associates the present model with the previous model.

$$\text{Acc} = \frac{\text{TP+TN}}{(\text{TP+TN+FP+FN})} \tag{3}$$

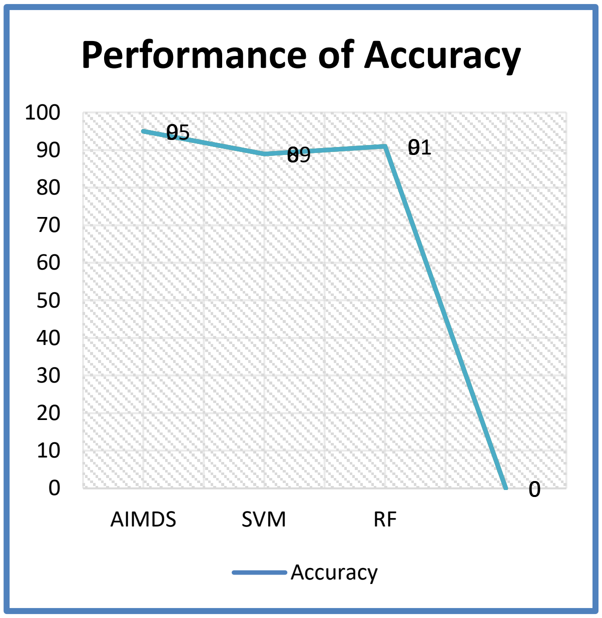


Fig. 4. Accuracy comparison.

D. Precision

Precision is the measurement of successfully recognized harmful users among all occurrences projected to be malicious [22]. It is determined as the ratio of TP (malicious users accurately predicted) to the total of TP and FP (malicious users wrongly categorized as non-malicious). A higher precision suggests that there are fewer false positives. Eq. (4) is used to evaluate the precision. Fig. 4 compares the present model with previous models.

$$\text{Prec} = \frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

E. Recall

The fraction of real malicious users properly recognized by the model is measured by Rc, also labelled as sensitivity or true positive rate. It is determined as the proportion of true positives to the total of TP and FN (malicious users categorized mistakenly as non-malicious). A better recall means that there are fewer false negatives. Eq. (5) is used to evaluate the recall. Fig. 5 compares the present model with previous models.

$$\text{Rc} = \frac{\text{TP}}{\text{TP+FN}} \tag{5}$$

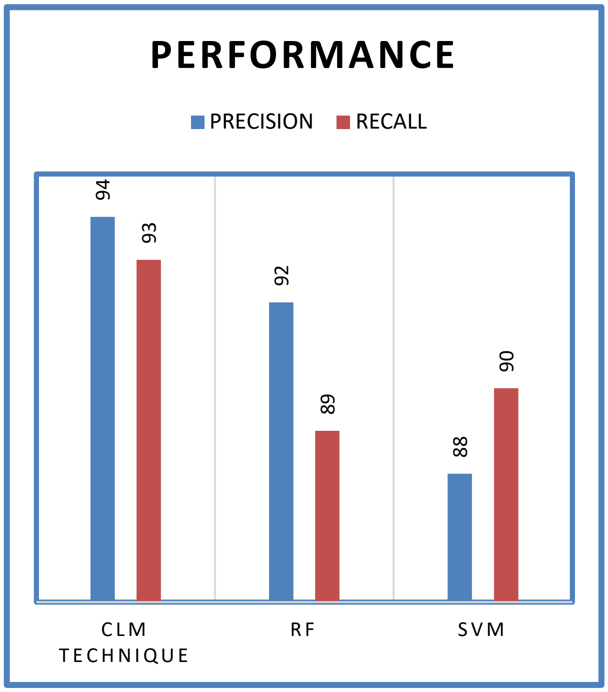


Fig. 5. Comparison of precision and recall.

F. F1 Score

The $f1_s$ combine accuracy and recall into a single statistic that balances their respective trade-offs. It provides an ample evaluation of the model's performance and is the harmonic mean of accuracy and recall. An increased F1-score suggests a better balance of accuracy and recall. The Eq. (6) evaluates the F1 score.

$$f1_s = \frac{2\times(\text{Prec}\times\text{Rc})}{(\text{Prec+Rc})} \tag{6}$$

Summarizing the values, the following Table I and Fig. 6 provide the overall performance of the CLM and optimization technique with the traditional algorithms. The experimental study of CLM and optimization technique used an amalgam of CNN, LSTM, and genetic algorithms (GA) to assess user behaviour in order to ascertain malevolent users. On the user behaviour dataset, which encompasses of user behaviour data gathered from a web platform, the performance of the suggested strategy was assessed. The outcomes show how well CLM technique (Convolution neural networks and LSTM) and optimization technique appropriately classify malicious users based on their behaviour patterns.

On the, which encompasses of TwiBot-20 dataset gathered from an online platform, the performance of the suggested strategy was assessed. The outcomes show how well CLM technique (Convolution neural networks and LSTM) and optimization technique perform in correctly classifying malicious users based on their behaviour patterns. Fig. 7 gives the portrayal of a comparison of evaluation metrics.

The proposed methodology consistently outperforms existing methodologies and traditional models, as demonstrated by the assessment measures. The genetic algorithm's ability to adapt is a crucial factor in accomplishing enhanced performance through hyper parameter optimization and feature selection. Conventional models may face difficulties in twigging the ever-changing and dynamic aspects of user behavior, while the proposed model excels in identifying intricate patterns.

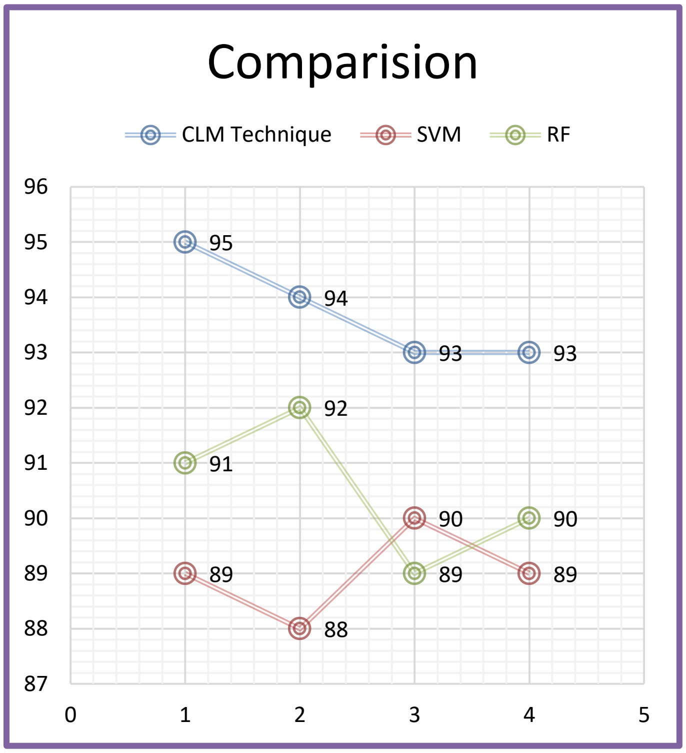


Fig. 6. Performance of proposed approach.

TABLE I. PERFORMANCE EVALUATION

Metric	Techniques		
	Proposed Technique	*SVM*	*RF*
Accuracy	95	89	91
Precision	94	88	92
Recall	93	90	89
F1 Score	93	89	90

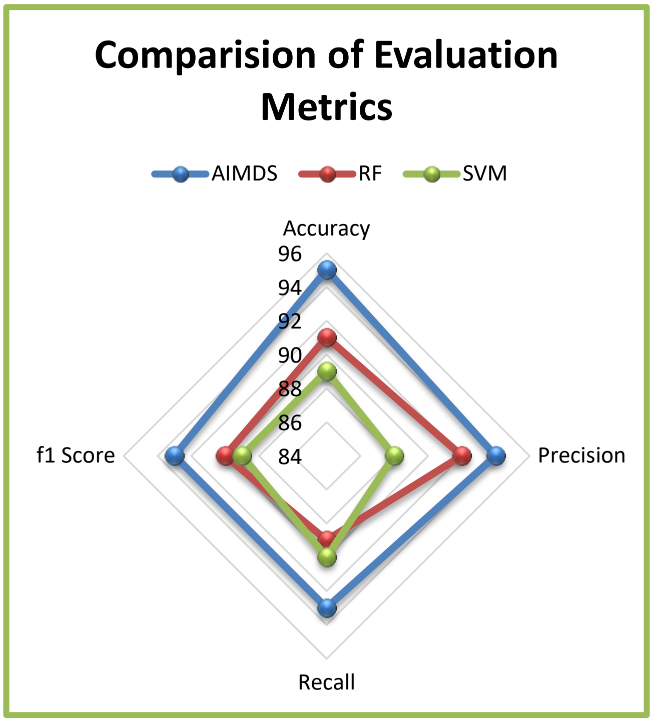


Fig. 7. Comparison of evaluation of metrics.

The proposed model that is anticipated has extraordinary performance; nonetheless, it is not immune to constraints. The efficacy of the model may differ contingent on the distinguishing features of various social platforms as well as the characteristics of malicious activities. Furthermore, the prospective for further research lies in exploring the interpretability of the model, explicitly addressing issues regarding the opaque nature of deep learning models.

V. Conclusion

The paper proposed a novel architecture a CLM technique (Convolution neural networks and LSTM) and an optimization technique to detect harmful user behavior using user behavior analysis in this study. In reliably detecting fraudulent users, an amalgam of CNN, LSTM networks, and genetic algorithms (GA) produced promising results. The model efficiently caught spatial patterns in the TwiBot-20 dataset by utilizing CNN. To capture temporal interdependence and sequential patterns in user behavior sequences, LSTM networks were used. The incorporation of genetic algorithms assisted in the optimization of model parameters and the improvement of model performance. On the TwiBot-20 dataset, the CLM and optimization technique surrogate conventional machine learning algorithms including SVM and Random Forest in terms of Acc, Prec, Rc, and $F1_s$.This demonstrates the utility of deep learning and genetic algorithms for identifying harmful user behavior. Overall, the technique proposed in this study provides a strong foundation for identifying fraudulent user behavior using deep learning methods. It paves the door for future research in deep learning, genetic algorithms, and user behavior analysis, paving the way for more advanced and accurate detection systems. Data on user behavior may not be sufficient to provide an ample portrait of harmful activity. The detection system's accuracy may be enhanced by additional data sources such as network traffic statistics, device information, and contextual data. To escalate the detection capacity, the future scope might investigate the integration of numerous data modalities. The prospect of detecting detrimental user behavior in online social networks is vast and promising. Ongoing exploration in these domains will not only enhance contemporary models but also aid in the conception of online security systems that are more ethical, transparent, and user-friendly. The initiative aims to tackle the complex issues presented by malicious user behavior in the digital domain by utilizing a multidisciplinary approach that encompasses computer science, social sciences, and ethics.

References

[1] Al-Hassan, M., Abu-Salih, B., & Al Hwaitat, A. (2023). DSpamOnto: An Ontology Modelling for Domain-Specific Social Spammers in Microblogging. *Big Data and Cognitive Computing*, *7*(2), 109.

[2] Han, R., Kim, K., Choi, B., & Jeong, Y. (2023). A Study on Detection of Malicious Behavior Based on Host Process Data Using Machine Learning. *Applied Sciences*, *13*(7), 4097.

[3] Hayawi, K., Saha, S., Masud, M. M., Mathew, S. S., & Kaosar, M. (2023). Social media bot detection with deep learning methods: a systematic review. *Neural Computing and Applications*, *35*(12), 8903-8918.

[4] El-Ghamry, A., Darwish, A., & Hassanien, A. E. (2023). An optimized CNN-based intrusion detection system for reducing risks in smart farming. *Internet of Things*, *22*, 100709.

[5] Alkahtani, H., & Aldhyani, T. H. (2022). Artificial intelligence algorithms for malware detection in android-operated mobile devices. *Sensors*, *22*(6), 2268.

[6] Ranjan, R., & Kumar, S. S. (2022). User behaviour analysis using data analytics and machine learning to predict malicious user versus legitimate user. *High-Confidence Computing*, *2*(1), 100034.

[7] Lazarov, A. D., & Petrova, P. (2022). Modelling activity of a malicious user in Computer Networks. *Cybernetics and information technologies*, *22*(2), 86-95.

[8] Jabar, T., Singh, M. M., & Al-Kadhimi, A. A. (2022, March). Mobile Advanced Persistent Threat Detection Using Device Behavior (SHOVEL) Framework. In *Proceedings of the 8th International Conference on Computational Science and Technology: ICCST 2021, Labuan, Malaysia, 28–29 August* (pp. 495-513). Singapore: Springer Singapore.

[9] Shen, X., Lv, W., Qiu, J., Kaur, A., Xiao, F., & Xia, F. (2022). Trust-aware detection of malicious users in dating social networks. *IEEE Transactions on Computational Social Systems*.

[10] Jain, A. K., Sahoo, S. R., & Kaubiyal, J. (2021). Online social networks security and privacy: comprehensive review and analysis. *Complex & Intelligent Systems*, *7*(5), 2157-2177.

[11] Senthil Raja, M., & Arun Raj, L. (2021). Detection of malicious profiles and protecting users in online social networks. *Wireless Personal Communications*, 1-18.

[12] Gururaj, H. L., Tanuja, U., Janhavi, V., & Ramesh, B. (2021). Detecting malicious users in the social networks using machine learning approach. *International Journal of Social Computing and Cyber-Physical Systems*, *2*(3), 229-243.

[13] Khaund, T., Kirdemir, B., Agarwal, N., Liu, H., & Morstatter, F. (2021). Social bots and their coordination during online campaigns: A survey. *IEEE Transactions on Computational Social Systems*, *9*(2), 530-545.

[14] Rahman, M. S., Halder, S., Uddin, M. A., & Acharjee, U. K. (2021). An efficient hybrid system for anomaly detection in social networks. *Cybersecurity*, *4*(1), 1-11.

[15] Sansonetti, G., Gasparetti, F., D'aniello, G., & Micarelli, A. (2020). Unreliable users detection in social media: Deep learning techniques for automatic detection. *IEEE Access*, *8*, 213154-213167.

[16] Terumalasetti, S. (2022, August). A Comprehensive Study on Review of AI Techniques to Provide Security in the Digital World. In *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)* (pp. 407-416). IEEE.

[17] Wu, X., Sun, Y. E., Du, Y., Xing, X., Gao, G., & Huang, H. (2020). An efficient malicious user detection mechanism for crowdsensing system. In *Wireless Algorithms, Systems, and Applications: 15th International Conference, WASA 2020, Qingdao, China, September 13–15, 2020, Proceedings, Part I 15* (pp. 507-519). Springer International Publishing.

[18] Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, *7*, 1-29.

[19] Wanda, P., Hiswati, M. E., & Jie, H. J. (2020). DeepOSN: Bringing deep learning as malicious detection scheme in online social network. *IAES International Journal of Artificial Intelligence*, *9*(1), 146.

[20] Mou, G., & Lee, K. (2020). Malicious bot detection in online social networks: arming handcrafted features with deep learning. In *Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6–9, 2020, Proceedings 12* (pp. 220-236). Springer International Publishing.

[21] Samokhvalov, D. I. (2020). Machine learning-based malicious users' detection in the VKontakte social network. *Труды института системного программирования РАН*, *32*(3), 109-117.

[22] Rabbani, M., Wang, Y. L., Khoshkangini, R., Jelodar, H., Zhao, R., & Hu, P. (2020). A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing. *Journal of Network and Computer Applications*, *151*, 102507.

[23] https://botometer.osome.iu.edu/bot-repository/datasets.html [Dataset].

[24] Kim, J., Park, M., Kim, H., Cho, S., & Kang, P. (2019). Insider threat detection based on user behavior modeling and anomaly detection algorithms. *Applied Sciences*, *9*(19), 4018.

[25] Qiu, J., Shen, X., Guo, Y., Yao, J., & Fang, R. (2019, August). Detecting malicious users in online dating application. In *2019 5th International Conference on Big Data Computing and Communications (BIGCOM)* (pp. 255-260). IEEE.

[26] Kiran, K., Manjunatha, C., Harini, T. S., Shenoy, P. D., & Venugopal, K. R. (2019, March). Identification of anomalous users in Twitter based on user behaviour using artificial neural networks. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). IEEE.

[27] Hong, T., Choi, C., & Shin, J. (2018). CNN-based malicious user detection in social networks. *Concurrency and Computation: Practice and Experience*, *30*(2), e4163.

[28] Yu, J., Wang, K., Li, P., Xia, R., Guo, S., & Guo, M. (2017). Efficient trustworthiness management for malicious user detection in big data collection. *IEEE Transactions on Big Data*, *8*(1), 99-112.

[29] Saracino, A., Sgandurra, D., Dini, G., & Martinelli, F. (2016). Madam: Effective and efficient behavior-based android malware detection and prevention. *IEEE Transactions on Dependable and Secure Computing*, *15*(1), 83-97.

[30] Khan, M. U. S., Ali, M., Abbas, A., Khan, S. U., & Zomaya, A. Y. (2016). Segregating spammers and unsolicited bloggers from genuine experts on twitter. *IEEE Transactions on Dependable and Secure Computing*, *15*(4), 551-560.

[31] Khan, M. U. S., Ali, M., Abbas, A., Khan, S. U., & Zomaya, A. Y. (2016). Segregating spammers and unsolicited bloggers from genuine experts on twitter. *IEEE Transactions on Dependable and Secure Computing*, *15*(4), 551-560.

Exploring a Novel Machine Learning Approach for Evaluating Parkinson's Disease, Duration, and Vitamin D Level

Md. Asraf Ali[1], Md. Kishor Morol[2], Muhammad F Mridha[3], Nafiz Fahad[4], Md Sadi Al Huda[5], Nasim Ahmed[6]
Department of Computer Science, American International University- Bangladesh, Dhaka, Bangladesh[1, 2, 3]
Computer Science Miscellaneous and Applications,
Artificial Intelligence Research and Innovation Lab (AIRIL), Dhaka, Bangladesh[4, 5]
School of Computer Science, The University of Sydney, Sydney, Australia[6]

Abstract—**Parkinson's disease is an increasingly prevalent, degenerative neurological condition predominantly afflicting individuals aged 50 and older. As global life expectancy continues to rise, the imperative for a deeper comprehension of factors influencing the course and intensity of PD becomes more pronounced. This investigation delves into these facets, scrutinizing various parameters including patient medical history, dietary practices, and vitamin D levels. A dataset comprising 50 PD patients and 50 healthy controls, sourced from Dhaka Medical Institute, serves as the foundation for this study. Machine learning techniques, notably the Modified Random Forest Classifier (MRFC), are harnessed to prognosticate both PD severity and duration. Strikingly, the MRFC-based prediction model for PD severity attains an impressive accuracy of 97.14%, while the predictive model for PD duration demonstrates an accuracy of 95.16%. Noteworthy is the observation that vitamin D levels are notably higher in the healthy cohort compared to PD-afflicted individuals, exerting a substantial positive influence on both the severity and duration predictions, surpassing the influence of other measured parameters. This inquiry underscores the practicality of machine learning in forecasting PD progression and duration and underscores the pivotal role of vitamin D levels as a predictive factor. These discoveries provide invaluable insights into advancing our comprehension and management of PD in an aging population.**

Keywords—Parkinson's disease; machine learning; vitamin D; severity; disease duration

I. Introduction

Parkinson's disease is a progressive neurological condition characterized by mobility restriction, primarily affecting the central nervous system. This is a consequence of the depletion of substantia nigra cells, which are responsible for producing dopamine, a critical factor in regulating movement [1]. This ailment manifests through motor and non-motor symptoms, often beginning subtly with hand tremors and gradually eroding motor control, affecting over 10 million individuals worldwide, particularly those aged 60 and above [2]. Understanding the early signs of Parkinson's disease (PD) is vital for effective intervention, but misconceptions and inadequate awareness persist, hindering timely diagnosis [3]. PD's impact on neural networks governing bodily movements is profound, involving critical brain regions such as the basal ganglia and the substantia nigra [4]. Typical PD symptoms encompass tremors, stiffness, slow movement, balance issues, and gait problems, alongside non-motor symptoms like depression, constipation, and sleep disturbances [5]. Recognizing these symptoms can be challenging due to their variability, emphasizing the importance of consulting neurophysicians, especially for older individuals [6].

Moreover, Parkinson's disease is on the rise in Bangladesh, largely due to increased life expectancy and a growing population. As vitamin D deficiency emerges as a potential risk factor for the condition, researchers are delving into the interplay among patient lifestyles, vitamin D levels, and the application of machine learning techniques to detect Parkinson's disease. Notably, there exists an inverse relationship between the severity of Parkinson's disease, its symptoms, and cognitive function with serum 25(OH)D levels. Generally, individuals with Parkinson's disease tend to exhibit lower vitamin D levels compared to their healthy counterparts [7, 8]. Additionally, studies have pointed to a significant occurrence of vitamin D deficiency among middle-aged women in Bangladesh [9]. Diverse machine learning methods, including Artificial Neural Networks (ANN), Decision Trees, and Support Vector Machines (SVM), have been employed to predict Parkinson's disease based on a variety of features [10]. Nevertheless, there remains an ongoing debate surrounding the link between vitamin D and Parkinson's disease, primarily due to disparities in the studied populations and methodological constraints [11][12].

Therefore, utilizing a custom model, this current study aims to shed light on the connection between PD severity and duration with vitamin D levels. It builds upon existing research to explore the potential of machine learning in predicting PD, offering a promising avenue for timely intervention. Key contributions of this paper are mentioned below:

1) *Utilized* supervised machine learning approaches to anticipate Parkinson's disease (PD) using a dataset comprising 50 PD patients and 50 healthy individuals from Bangladesh.

2) *Collected* demographic data and clinical features of PD participants, including age, education, disease duration, cardinal features of the disease, and disease severity, to build a comprehensive dataset.

3) The vitamin D levels of participants were evaluated, and vitamin D deficiency was classified as values below 30ng/mL.

4) Investigated correlations among various PD features, including positive and negative correlations.

5) Employed several machine learning algorithms to predict PD based on the dataset, including Random Forest, Decision Tree, Naive Bayes, Logistic Regression, Nearest Neighbor (KNN), and a Modified Random Forest Classifier (MRFC).

6) Customize MRFC model with specific settings, including the number of trees, maximum depth, maximum number of features, verbosity, mode, warm start, and thread usage, to optimize its performance.

7) Introduced performance evaluation metrics such as sensitivity (Sn) and specificity (Sp) into the confusion matrix to assess the classification results.

Therefore, the rest of the paper is designed as follows: Section II is literature review, Section III is about materials and methods, Section IV delves into results and discussion, Section V is Discussion, and in Section VI conclusion is presented.

II. Literature Review

Various researchers do research in the field of Parkinson's disease detection. Quan et al. (2021) focus on exploiting dynamic speech features to identify voice alterations in patients with Parkinson's disease (PD). Using a Bidirectional LSTM model instead of the static features seen in conventional machine learning models, it greatly increases the accuracy of PD identification. The study makes use of a mixed-gender database from GYENNO SCIENCE Parkinson that has 45 patients (15 Healthy Controls, 30 PD cases). However, information regarding feature engineering and data preprocessing is missing. According to the experimental results, the suggested Bidirectional LSTM model has an accuracy of 75.56% [13]. Shinde et al. (2019) improved the diagnosis of Parkinson's disease (PD) by using convolutional neural networks (CNNs) with neuromelanin-sensitive magnetic resonance imaging (NMS-MRI), a computer-based method. This strategy achieves better results than current approaches: 80% of tests correctly identify PD from healthy controls, and 85.7% correctly identify PD from atypical Parkinsonian disorders. It detects minute alterations in the substantia nigra pars compacta (SNc) by using CNNs for feature extraction and data augmentation. Sensitivity, specificity, and ROC curves are some of the evaluation criteria that demonstrate its excellent performance in the diagnosis and categorization of Parkinson's disease [14]. Noor et al. (2020) investigated the use of deep learning to identify neurological conditions using several MRI modalities, with a focus on schizophrenia, Parkinson's disease, and Alzheimer's disease. Convolutional Neural Networks (CNNs) are emphasized as being better at detecting these illnesses in its overview of current deep learning techniques. Datasets including MIRIAD, Open fMRI, PPMI, ADNI, COBRE, fastMRI, and FBIRN are used in the work along with discussion of problems and recommendations for future research areas. There is no specific discussion of feature extraction techniques or data preprocessing. The emphasis is on deep learning techniques, with CNNs being identified as the most effective way [15].

III. Materials and Methods

Numerous researchers in various disciplines have recently adopted machine learning-based approaches to get better accuracy from the analysis of complicated data [13]. The main objective of this study was to anticipate PD through various characteristics data which is complicated. Thus, supervised machine learning approaches were used to achieve the objectives of this study.

A. Dataset Description

To conduct this research, we collected a dataset of 50 PD patients and 50 healthy people from Sir Salimullah Medical College, where all subjects were Bangladeshi. Demographic Data and Clinical Features of PD participants are presented in Table I. The clinical states of the recruited subjects were evaluated using the Movement Disorder Society-Unified Parkinson's Disease Rating Scale [16], and the severity of Parkinson's disease was determined according to the study's criteria [17]. In the laboratory of Sir Salimullah Medical College, serum 25-hydroxyvitamin D levels were determined in both the PD and healthy groups using CLIA kits using the radioimmunoassay technique. In the present investigation, blood vitamin D levels of 30ng/mL were judged normal, whereas values below 30ng/mL were deemed insufficient. These include age, sex, education, occupation, socioeconomic status, nature of work, cardinal features, disease duration, disease severity, smoking history, caffeine history, and vitamin D level. Additionally, the age comparison between PD and healthy participants is presented in Fig. 1, where PD participants are younger than healthy participants.

TABLE I. The Demographic Data and Clinical Features of PD Participants

Variables		Case	
		N	Percentage
Age	<60	1	2
	≥ 60	49	98
Education	Diploma	31	62
	Illiterate	11	22
	Under Diploma	8	16
Disease Duration	less than 5 years	26	52
	5 to 10 years	17	34
	Greater than 10	7	14
Cardinal Features of Disease	tremor	30	60
	rigidity	5	10
	bradykinesia	12	24
	postural instability	3	6
Disease Severity	1-2.5	26	52
	2.5-3	18	36
	>3	6	12

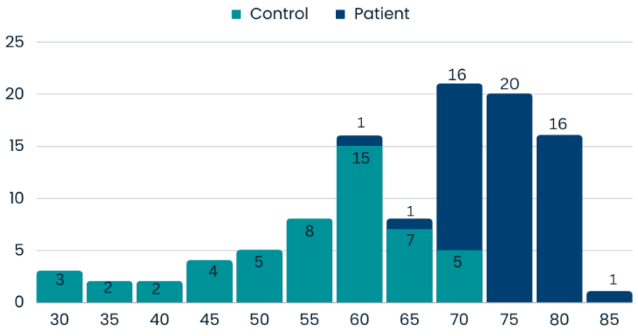


Fig. 1. Age comparison between PD and healthy participants.

Moreover, various features are causes of PD. The correlations among the features of PD are shown in Fig. 2. There are two types of correlations among the features - positive & negative. A positive correlation indicates that if the characteristic improves, the linked feature likewise increases, and if the feature falls, the connected feature reduces as well. Both aspects move in simultaneously, and their connection is linear. A negative correlation indicates that if one property's value rises, the linked trait's value falls, and vice versa.

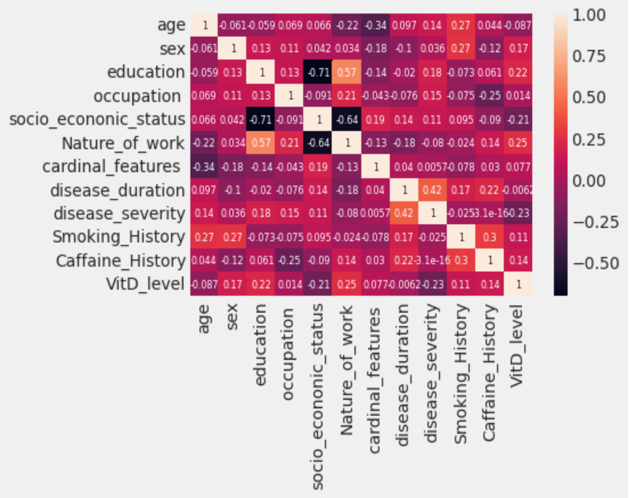


Fig. 2. Correlations among the features.

B. Model Description

1) Random forest classifier: A well-known supervised machine learning algorithm is the Random Forest. It finds application in both classification and regression challenges within machine learning. The Random Forest serves as a classifier that enhances the predictive accuracy of a dataset by averaging the outcomes of multiple decision trees generated on different subsets of the dataset. When applied to address regression problems, the Random Forest Algorithm utilizes the mean squared error (MSE) from each node's data branch [17]. However, the formula is mentioned below in the equation number i.

$$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \quad (\mathrm{fi} - \mathrm{yi})^2 \qquad (1)$$

In this context, N denotes the number of data points, fi signifies the model's output, and yi signifies the actual value for the specific data point i. The formula is employed to calculate the distance from the forecasted value to each node, and this outcome is pivotal in establishing the most suitable path through the forest.

2) Decision tree: The decision tree is a method for supervised learning that can effectively address classification and regression problems, primarily in solving classification tasks [18]. The terminology implies its utilization of a tree-like diagram to present predictions derived from a sequence of feature-driven divisions. It commences with a central node and culminates in a determination made at the terminal leaf. Therefore, in Fig. 3, a decision tree is mentioned [19].

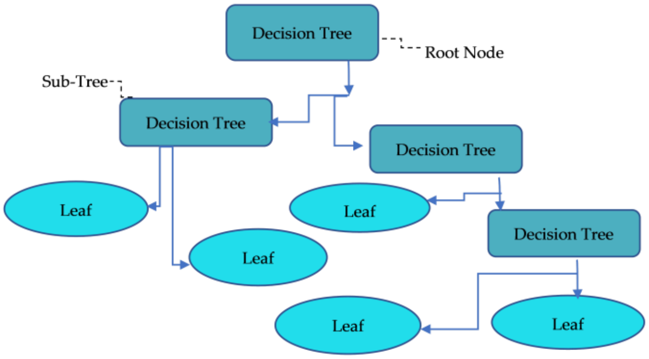


Fig. 3. Decision tree.

The decision trees utilize a representation known as the Sum of Products (SOP). The term Disjunctive Normal Form (DNF) is an alternative designation for the Sum of Products (SOP). In this context, each branch culminating in a node of the same category represents a conjunction (product) of values for that category. Conversely, separate branches ending in the same category constitute a disjunction (sum).

3) Naive bayes: The Naive Bayes algorithm is an algorithm for supervised learning. The Naive Bayes classifier is based on the conditional probability principle. It is straightforward and quick to predict the category of the test data set. It is also effective for multiclass prediction [17]. When the independence assumption holds true, the Naive Bayes classifier outperforms other models. The equation of Naïve Bayes is mentioned below in equation number ii.

$$\mathrm{P(A \mid B)} = \frac{\mathrm{P(B \mid A).P(A)}}{\mathrm{P(B)}} \qquad (2)$$

The posterior probability P (c|x) can be computed using the prior probabilities P(c), P(x), and P(x|c), thanks to Bayes' theorem [19]. Naive Bayes classifiers assume that the impact of one predictor value (x) on a specific class (c) is independent of the impact of other predictor value combinations. Here, conditional independence on class labels is granted.

4) Logistic regression: Logistic regression is an example of supervised learning. It is used to solve classification problems, with the most common application being binary logistic regression with a binary outcome. Number iii equations underlying logistic regression:

$$y = \frac{e^{(b0+b1\times x)}}{1+e^{(b0+b1\times x)}} \qquad (3)$$

where, y represents the predicted output, b0 represents the bias or intercept term, and b1 represents the single input value (x) coefficient. Each input data column contains a b coefficient (a constant real value) that must be learned from your training data [18].

5) K-Nearest Neighbor (KNN): K-Nearest Neighbors (KNN) is a relatively straightforward supervised machine learning technique. In this method, new data or instances are categorized based on their similarity to the existing ones. KNN follows a "lazy learning" approach, setting it apart from the previously discussed classifiers. It doesn't actively progress during training and primarily involves storing the training data. Its classification efforts only come into play when fresh, unlabeled data emerges [19].

KNN demonstrates the optimal performance when the nearest neighbors to a data point all belong to the same category. The underlying assumption is that if all nearby neighbors concur, a new data point will likely fall within the same group. Two compelling reasons for employing KNN are its simplicity in terms of comprehension and application. However, KNN's accuracy hinges on the chosen distance metric, and under certain distance metrics, it can achieve 100 percent accuracy [18].

Nevertheless, the computational cost of finding the closest neighbors of KNN on vast datasets can be quite significant. Furthermore, noisy data can potentially disrupt KNN classification. Features with a wider range of values may dominate the distance metric, necessitating the normalization or scaling of features. Notably, due to its "lazy learning" approach, KNN often demands more substantial storage resources than eager classifiers. Thus, the effectiveness of KNN is closely tied to selecting an appropriate distance metric [17] [20].

6) Modified Random Forest Classifier (MRFC): This research utilized a modified Random Forest classifier (MRFC) as the proposed model (see Fig. 4). We choose MRFC for these reasons-

- The MRFC model with specific settings, including the number of trees, maximum depth, maximum number of features, verbosity, mode, warm start, and thread usage, to optimize its performance.
- Customizations are designed to fine-tune the Random Forest algorithm for the specific task of predicting Parkinson's disease severity and duration, aiming to achieve better accuracy and performance on this given dataset.

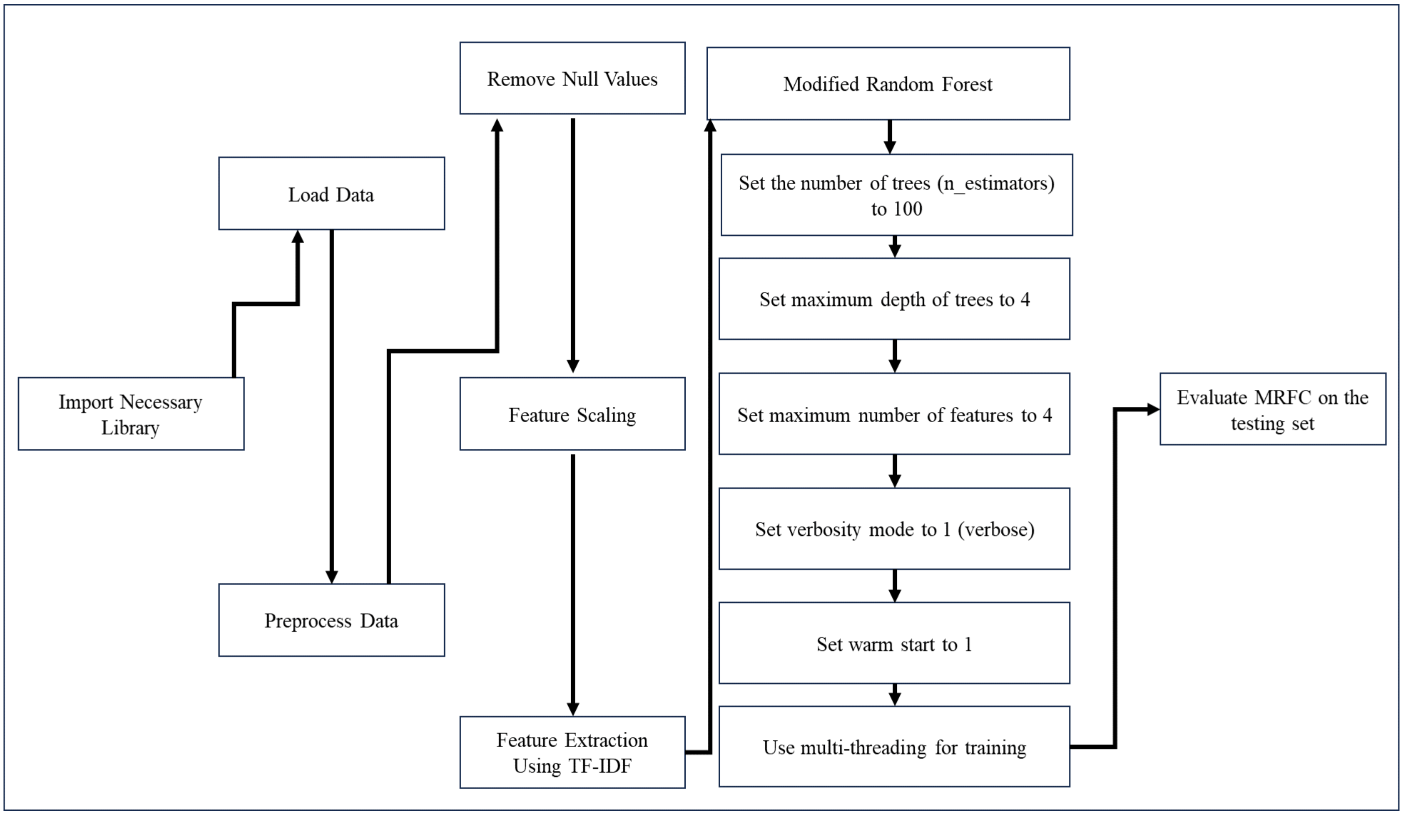

Remove Null Values
Modified Random Forest
Load Data
Set the number of trees (n_estimators) to 100
Set maximum depth of trees to 4
Import Necessary Library
Feature Scaling
Set maximum number of features to 4
Evaluate MRFC on the testing set
Set verbosity mode to 1 (verbose)
Preprocess Data
Set warm start to 1
Feature Extraction Using TF-IDF
Use multi-threading for training

Fig. 4. Modified random forest classifier.

- Parameter values such as a tree count of 100, a maximum depth of 4, and a maximum number of features of 4 were utilized to adjust the random forest classifier and attain the best precision level.
- The MRFC was instrumental in achieving the best accuracy on the dataset.

Therefore, MRFC is chosen for achieving the best accuracy than the existing models. However, MRFC was instrumental in achieving the best accuracy on this dataset. Thirty percent of the data were allocated for testing, with the remaining 70% for training. The random forest classifier was adjusted to attain the best precision level. Various parameter values were utilized, such as a tree count of 100 (n estimators = 100), a maximum depth of 4, and a maximum number of features of 4. The parameter "verbose" was set to 1, representing a moderate level of verbosity, and "warm start" was also set to 1.

The training process for the model utilized multiple threads to enhance efficiency. The use of multi-threading varies among different learning algorithms, and if not specified, the number of threads will be set to the number of cores in the system, with a maximum of 32. Setting the number of threads significantly higher than the number of processors can considerably slow down the training time. Future work may consider altering the default values to further enhance the model's performance. The improved model provides more accurate estimates of the severity and duration of PD.Accurate estimates of the severity and duration of PD.

In Fig. 4, the Modified Random Forest Classifier (MRFC) is customized or modified in several key aspects compared to the standard Random Forest classifier:

a) Number of Trees (n_estimators): MRFC uses a specific value for the number of trees in the forest. In the description, it mentions a tree count of 100 (n_estimators=100). This is a customization because the number of trees in a Random Forest can vary, and choosing the appropriate number can impact the model performance.

b) Maximum Depth and Maximum Number of Features: MRFC specifies the maximum depth of the trees and the maximum number of features considered at each split. In the description, it mentions a maximum depth of 4 and a maximum number of features of 4. These values determine the complexity of individual trees in the forest.

c) Verbosity Mode (Verbose): MRFC introduces a verbosity model to control the level of detail in the model's output. It mentions setting the verbosity level to 1 for small details. This customization helps in better understanding the model's behavior and performance.

d) Warm Start: MRFC mentions using a warm start level of 1. Warm start is a technique where the previously trained forest is used as an initialization for the next forest. This can speed up training and potentially improve convergence.

e) Thread Usage (num_threads): MRFC controls the number of threads used during training. It mentions that multi-threading is used and that the number of threads (num_threads) is set based on the available system resources. This customization optimizes the training process for efficiency.

These customizations are designed to fine-tune the Random Forest algorithm for predicting Parkinson's disease severity and duration, aiming to achieve better accuracy and performance on the given dataset. Customization of Random Forest parameters is a common practice in machine learning to adapt the model to the characteristics of the data and the problem at hand.

7) Performance evaluation matrix: One common tool for assessing a solution to a classification problem is the confusion matrix. The method is flexible enough to be used for both binary and multiclass classification problems. Confusion matrices show which values were correctly classified as TP, which were incorrectly classified as FP, which were incorrectly classified as FN, and which were incorrectly classified as TN. Sensitivity (Sn) and specificity (Sp) are the most popular performance metrics for classifying based on these values [20] [21] [22] [23] [24]. Using the confusion matrix values, these metrics are computed using Eq. (4) and Eq. (5).

$$\text{Sensitivity (Sn)} = (TP)/\ (TP + FN) \qquad (4)$$

$$\text{Specificity (Sp)} = \ TN/\ (TN + FP) \qquad (5)$$

Here, TP means true positive, TN means true negatives, and FP means false positives.

IV. Results

Usually, the dataset allows us to discover the characteristics that affect Parkinson's disease. In this section, we present progressively the performance of various machine learning models using our dataset for detecting PD severity and its duration, and the various factors that are the causes of PD including age, gender, education, vitamin D level, etc. The performance evaluation for PD severity and its duration of various machine learning models are presented in Table II. The higher accuracy score was found using the Modified Random Forest Classifier model compared to other machine learning models that were applied in this study for detecting PD severity ((97.14%), and its duration (95.16%).

TABLE II. Model Performance Evaluation for Severity & Duration

Model	Accuracy Score (Severity)	Accuracy Score (Duration)
Modified Random Forest Classifier (MRFC)	97.14%	95.16%
Decision Tree	96%	92.57%
Random Forest Classifier	87.43%	89.26%
Gaussian NB	86.67%	68.57%
Logistic Regression	81.08%	83.33%
KNN	75%	74.07%
Linear SVC	74.29%	80.66%

A. Parkinson Severity

The performance of different classification models on severity detection using confusion matrix are standard, which are shown in Fig. 5. Then, we found that the most significant influence on PD severity is played by the feature of vitamin D, and then the next three features are age, disease duration, and occupation (see Fig. 6).

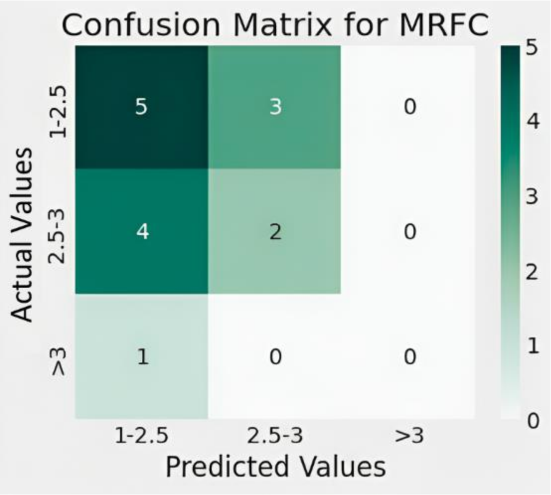


Fig. 5. Confusion matrix results to assess the performance of different classification models on severity detection.

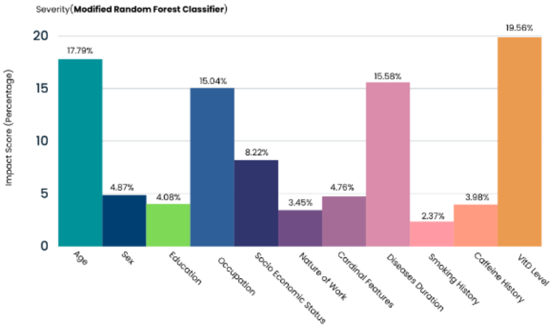


Fig. 6. Impact of different features for severity detection in PD patients.

B. Parkinson Duration

The performance of different classification models on duration detection using a confusion matrix is standard, which is shown in Fig. 7. Then, we found that the vitamin D feature significantly influences PD severity, and the following three features are disease duration, age, and occupation (see Fig. 8).

C. Vitamin D Level

The characteristic that has the most significant influence, as in this research, is vitamin D. We have conducted further research on the vitamin D levels of PD and healthy subjects, allowing us to define this characteristic of Parkinson's disease more clearly.

The comparison of vitamin D levels (see Fig. 9) between healthy individuals and patients has made the impact of vitamin D on Bangladeshi citizens quite obvious. The rate of healthy individuals, who have an average age of almost 50 and do their normal employment outside the household, have stronger vitamin D compared to the PD patient, who seldom goes outdoors since Bangladesh is one of the warmer regions in the world and sunlight can be found virtually every day.

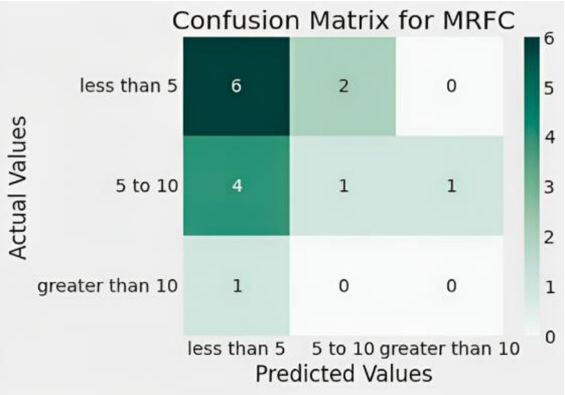


Fig. 7. Confusion matrix results to assess the performance of different classification models on duration detection.

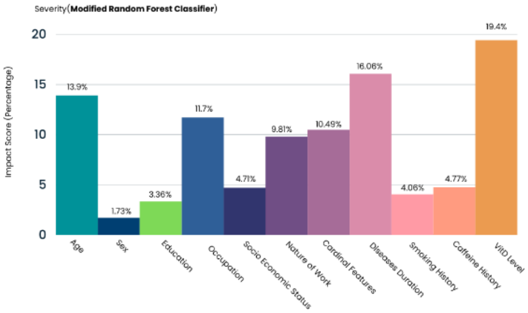


Fig. 8. Impact of different features for duration detection in PD patients.

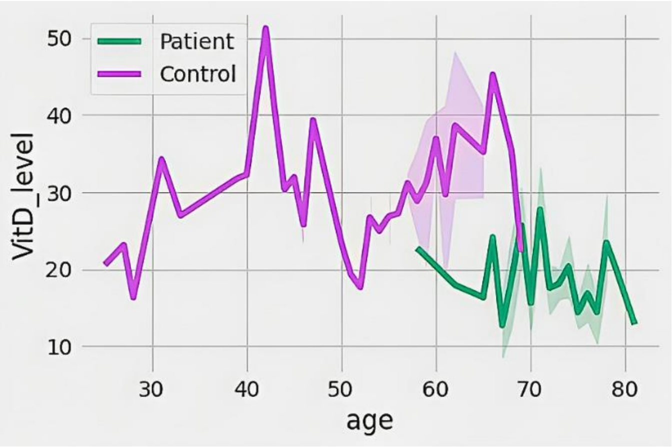


Fig. 9. Age-dependent variation in vitamin D levels in Parkinson's disease and healthy individuals.

V. Discussion

In this study, the Modified Random Forest Classifier (MRFC) demonstrated remarkable predictive capabilities, achieving impressive accuracies of 97.14% for Parkinson's disease (PD) severity and 95.16% for its duration. The selection of MRFC was based on its ability to yield the highest accuracy on the dataset. The Random Forest algorithm underwent customizations, including parameter fine-tuning, verbosity level set to 1, and the implementation of a warm start technique, tailored to the data's characteristics to enhance overall performance. Despite considering various parameters such as patient medical history and dietary practices, the significant positive impact of vitamin D levels on both severity and duration predictions outweighed other influences. The result is an improved model providing more accurate estimates of PD severity and duration.

Therefore, Customizations of the Random Forest algorithm were made to fine-tune it for predicting Parkinson's disease severity and duration, aiming to achieve better accuracy and performance on the given dataset. Customizing Random Forest parameters is a common practice in machine learning to adapt the model to the characteristics of the data and the problem at hand. The performance of different classification models on severity detection using a confusion matrix is standard practice. The most significant influence on Parkinson's disease severity is played by the feature of vitamin D, followed by age, disease duration, and occupation. Supervised machine learning approaches were used to anticipate Parkinson's disease through various characteristics data.

VI. Conclusion

Parkinson's disease remains a complex neurodegenerative condition influenced by various factors, including genetic predisposition and lifestyle variables. This study has explored the role of vitamin D as a potential predictor of PD severity and duration, shedding light on its significance in comprehending the disease's progression. The observed disparities in vitamin D levels between healthy individuals and PD patients and its age-dependent impact on disease severity and duration emphasize the need for further investigation. Nevertheless, it is crucial to acknowledge the limitations of our dataset, especially the absence of severity and duration data for the healthy group and the relatively small sample size. Future research efforts should expand the dataset to encompass a broader spectrum of predictive factors and their interactions to enhance our understanding of PD development and progression. This ongoing exploration holds the potential to advance our knowledge and lay the groundwork for more effective interventions and management strategies for Parkinson's disease. This study lies in examining vitamin D's role as a predictor, it serves as a starting point for more comprehensive research in this domain, compared with other existing work, and uses other datasets, which may ultimately lead to improved approaches for addressing this challenging condition.

Acknowledgment

The authors express their gratitude to the Artificial Intelligence Research and Innovation Lab (AIRIL) for their supporting efforts in conducting this research. Additionally, the authors acknowledge and appreciate the support from the American International University-Bangladesh for covering the Article Processing Charges (APC).

References

[1] Khan, S. S., Janrao, S., Srivastava, S., Singh, S. B., Vora, L., & Khatri, D. K. (2023). GSK-3β: An Exuberating Neuroinflammatory Mediator in Parkinson's Disease. *Biochemical Pharmacology*, 115496.

[2] Reeve, A., Simcox, E., & Turnbull, D. (2014). Ageing and Parkinson's disease: why is advancing age the biggest risk factor?. *Ageing research reviews*, *14*, 19-30

[3] Swallow, D. M., & Counsell, C. E. (2023). The evolution of diagnosis from symptom onset to death in progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) compared to Parkinson's disease (PD). Journal of Neurology, 270(7), 3464-3474.

[4] Takakusaki, K. (2017). Functional neuroanatomy for posture and gait control. *Journal of movement disorders*, *10*(1), 1.

[5] Chiew, A., Mathew, D., Kumar, C. M., Seet, E., Imani, F., & Khademi, S. H. (2023). Anesthetic Considerations for Cataract Surgery in Patients with Parkinson's Disease: A Narrative Review. *Anesthesiology and Pain Medicine*, *13*(3).

[6] Goldenberg, M. M. (2008). Medical management of Parkinson's disease. *Pharmacy and Therapeutics*, *33*(10), 590.

[7] Zhang, H. J., Zhang, J. R., Mao, C. J., Li, K., Wang, F., Chen, J., & Liu, C. F. (2019). Relationship between 25‐Hydroxyvitamin D, bone density, and Parkinson's disease symptoms. *Acta Neurologica Scandinavica*, *140*(4), 274-280.

[8] Suzuki, M., Yoshioka, M., Hashimoto, M., Murakami, M., Kawasaki, K., Noya, M., ... & Urashima, M. (2012). 25‐hydroxyvitamin D, vitamin D receptor gene polymorphisms, and severity of Parkinson's disease. *Movement Disorders*, *27*(2), 264-271.

[9] Roth, D. E., Shah, M. R., Black, R. E., & Baqui, A. H. (2010). Vitamin D status of infants in northeastern rural Bangladesh: preliminary observations and a review of potential determinants. *Journal of health, population, and nutrition*, *28*(5), 458.

[10] Bind, S., Tiwari, A. K., Sahani, A. K., Koulibaly, P., Nobili, F., Pagani, M., & Tatsch, K. (2015). A survey of machine learning based approaches for Parkinson's disease prediction. *Int. J. Comput. Sci. Inf. Technol*, *6*(2), 1648-1655.

[11] Cunningham, P., & Delany, S. J. (2021). k-Nearest neighbour classifiers-A Tutorial. *ACM computing surveys (CSUR)*, *54*(6), 1-25.

[12] Modak, G., Das, S. S., Miraj, M. A. I., & Morol, M. K. (2022, March). A Deep Learning Framework to Reconstruct Face under Mask. In *2022, 7th International Conference on Data Science and Machine Learning Applications (CDMA)* (pp. 200-205). IEEE.

[13] Quan, C., Ren, K., & Luo, Z. (2021). A deep learning based method for Parkinson's disease detection using dynamic features of speech. *IEEE Access*, *9*, 10239-10252.

[14] Shinde, S., Prasad, S., Saboo, Y., Kaushick, R., Saini, J., Pal, P. K., & Ingalhalikar, M. (2019). Predictive markers for Parkinson's disease using deep neural nets with neuromelanin sensitive MRI. *NeuroImage: Clinical*, *22*, 101748.

[15] Noor, M. B. T., Zenia, N. Z., Kaiser, M. S., Mamun, S. A., & Mahmud, M. (2020). Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease, and schizophrenia. *Brain informatics*, *7*, 1-21.

[16] Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez‐Martin, P., ... & LaPelle, N. (2008). Movement Disorder Society‐sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS‐UPDRS): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, *23*(15), 2129-2170.

[17] Newberry, S. J., Chung, M., Shekelle, P., Booth, M., Liu, J., Maher, A., ... & Balk, E. (2016). Issues in the Assessment of Vitamin D Status for Clinical, Research, and Public Health Purposes. *Advances in Nutrition*, *7*(1), 49A-49A.

[18] Pérez-Castrillón, J. L., Dueñas-Laita, A., Gómez-Alonso, C., Jódar, E., del Pino-Montes, J., Brandi, M. L., & Chinchilla, S. P. (2023). Long-Term Treatment and Effect of Discontinuation of Calcifediol in Postmenopausal Women with Vitamin D Deficiency: A Randomized Trial. *Journal of Bone and Mineral Research*, *38*(4), 471-479.

[19] Fahad, N., Goh, K. O. M., Hossen, M. I., Tee, C., & Ali, M. A. (2023). Building a Fortress Against Fake News: Harnessing the Power of Subfields in Artificial Intelligence. *Journal of Telecommunications and the Digital Economy*, *11*(3), 68-83.

[20] Fahad, N., Goh, K. M., Hossen, M. I., Shopnil, K. S., Mitu, I. J., Alif, M. A. H., & Tee, C. (2023). Stand up Against Bad Intended News: An Approach to Detect Fake News using Machine Learning. *Emerging Science Journal*, *7*(4), 1247-1259.

[21] Sikandar, T., Rahman, S. M., Islam, D., Ali, M. A., Mamun, M. A. A., Rabbi, M. F., ... & Ahamed, N. U. (2022). Walking Speed Classification from Marker-Free Video Images in Two-Dimension Using Optimum Data and a Deep Learning Method. *Bioengineering*, *9*(11), 715.

[22] Sumathi, M. R., & Poorna, B. (2016). Prediction of mental health problems among children using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, *7*(1).

[23] Ali, W. (2017). Phishing website detection based on supervised machine learning with wrapper feature selection. *International Journal of Advanced Computer Science and Applications*, *8*(9).

[24] Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. *International Journal of Advanced Computer Science and Applications*, *10*(6)

Applying Big Data Analysis and Machine Learning Approaches for Optimal Production Management

Sarsenkul Tileubay[1], Bayanali Doszhanov[2], Bulgyn Mailykhanova[3],
Nurlan Kulmurzayev[4], Aisanim Sarsenbayeva[5], Zhadyra Akanova[6], Sveta Toxanova[7]
Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan[1,2,4,7]
Satbayev University, Almaty, Kazakhstan[3]
Kazakh National Pedagogical University, Almaty, Kazakhstan[5]
NARXOZ University, Almaty, Kazakhstan[6]

***Abstract*—In this research paper, we delve into the transformative potential of integrating Big Data analytics with machine learning (ML) techniques, orchestrating a paradigm shift in production management methodologies. Traditional production systems, often marred by inefficiencies stemming from data opacity, have encountered bottlenecks that throttle scalability and adaptability, particularly in complex, fluctuating markets. By harnessing the voluminous streams of data—both structured and unstructured—generated in contemporary production environments, and subjecting these data lakes to advanced ML algorithms, we unveil profound insights and predictive patterns that remain elusive under conventional analytical methods. Our discourse juxtaposes the multidimensionality of Big Data—emphasizing velocity, variety, veracity, and volume—with the finesse of ML models, such as neural networks and reinforcement learning, which adapt iteratively to the dynamism inherent in production landscapes. This symbiosis underpins a more holistic, anticipatory decision-making process, empowering stakeholders to pinpoint and mitigate operational hiccups, optimize supply chain vectors, and streamline quality assurance protocols, thereby catalyzing a more resilient, responsive, and cost-effective production framework. Furthermore, we explore the ethical contours of data stewardship in this context, advocating for a judicious balance between technological ascendancy and responsible data governance. The culmination of this exploration is the conceptualization of a predictive, self-regulating production ecosystem that thrives on continuous learning and improvement, dynamically calibrating itself in response to an ever-evolving market tableau and thereby heralding a new era of optimal, sustainable, and intelligent production management.**

Keywords*—Optimal production; smart manufacturing; machine learning; big data; management

I. Introduction

The transformative intersection of Big Data and machine learning (ML) represents a pioneering frontier in the realm of production management, poised to redefine traditional methodologies and infrastructures [1]. This integration marks a critical phase in the evolution of what's popularly known as Industry 4.0, where digitalization and intelligent analytics become the cornerstone of industrial operations [2]. Traditional production management strategies, although reliable over past decades, now face significant hurdles, primarily due to their limitations in handling the sheer volume and complexity of contemporary data and the dynamic nature of global markets [3].

The concept of Big Data is not new; however, its application within the industrial sector unveils new opportunities and challenges. Big Data refers to the enormous volume of data that inundates businesses daily and the corresponding analytics processes that seek to make sense of these data in various formats [4]. The characteristics of Big Data, often described by the four Vs (volume, velocity, variety, and veracity), suggest both the scale of data to be processed and the complexity involved in these operations [5]. Nevertheless, as some researchers articulate, the integration of Big Data into production systems is not merely a matter of handling large data volumes; it involves extracting actionable insights that can drive efficiency and innovation in production management [1].

Parallel to the Big Data revolution, machine learning has emerged as a powerful tool capable of providing sophisticated analyses and predictive insights in complex environments. ML algorithms, a subset of artificial intelligence, are designed to learn and improve from experience without being explicitly programmed, making them ideal for environments where data influx is continuous and variable [6]. Next studies have demonstrated ML's efficacy in enhancing various production aspects, including predictive maintenance and quality assurance, by allowing for more nuanced, data-driven decision-making processes [7, 8].

The fusion of Big Data and ML in production management necessitates a significant overhaul of existing infrastructures, necessitating substantial investments in both digital tools and human expertise [9]. Additionally, with the increased digitization of production data, issues surrounding cybersecurity and data privacy have come to the forefront, calling for robust security protocols and ethical data management practices [10, 11]. Some authors have emphasized the criticality of these aspects, advocating for a balanced approach between technological advancements and regulatory compliance [12, 13].

The potential benefits of integrating Big Data and ML into production management are substantial, extending beyond mere efficiency gains. This synergy is anticipated to engender more adaptable, resilient, and intelligent production systems, capable of predictive problem-solving and optimized resource management, thus delivering products and services that meet evolving market demands [14]. Through detailed case studies and practical evaluations, next studies have documented significant improvements in supply chain management, operational efficiency, and energy savings, attributing these advancements to the strategic leverage of Big Data and ML [15, 16].

This research paper, therefore, seeks to elaborate on the potential of Big Data and machine learning as a combined force reshaping production management. It aims to navigate through the theoretical discourse, practical challenges, and ethical considerations, drawing on contemporary studies and industrial applications to present a holistic view of this technological convergence. The goal is not only to highlight the transformative power of these technologies but also to identify pathways through which industries can navigate the complexities of integration, leveraging these tools for a more sustainable, efficient, and innovative production future.

II. Related Works

The scholarly landscape exploring the integration of Big Data and machine learning (ML) in production management is both vast and multidimensional, reflecting diverse methodologies, case studies, and theoretical analyses. This comprehensive review critically examines the pivotal literature in this domain, highlighting key findings, innovative approaches, and foundational theories that contribute to understanding this technological amalgamation's transformative potential.

A. *Big Data in Production Management*

Big Data's infusion into production landscapes has been revolutionary, with researchers highlighting its capacity to drive operational transparency and optimization. Li et al. (2022) presented one of the foundational frameworks for integrating Big Data analytics into manufacturing, emphasizing its role in real-time decision-making and efficiency enhancement through predictive insights [17]. Further expanding this discourse, a study by Tseng et al. (2021) introduced the concept of 'cyber-physical production systems,' illustrating how Big Data facilitates the digital synchronization of physical production activities, significantly enhancing operational agility and responsiveness [18].

B. *Evolution of Machine Learning in Industrial Applications*

The literature vividly documents ML's ascension in industrial environments, driven by its capacity for predictive accuracy and automation. An influential study by Qi et al. (2023) underscored ML's transformative effects in production settings, particularly highlighting its proficiency in streamlining production workflows through intelligent automation [19]. In a similar context, Ming et al. (2023) explored ML's implications for quality control, revealing how machine learning models outperform traditional statistical methods in identifying manufacturing defects, thereby ensuring higher product quality standards [20].

C. *Confluence of Big Data and Machine Learning*

The scholarly pursuit to harness Big Data and ML's combined capabilities has given rise to innovative paradigms in production management. Notably, Wang et al. (2023) provided groundbreaking insights by demonstrating how ML algorithms, when fed with diverse and extensive industrial Big Data, could predict production bottlenecks, thereby informing better resource allocation strategies [21]. Further, Serey et al. (2023) conducted an empirical analysis across various manufacturing sectors, revealing that companies employing Big Data-driven ML strategies witnessed substantial improvements in production scalability and customization [22].

D. *Ethical and Security Considerations*

The ethical and security dimensions of implementing Big Data and ML have been rigorously debated within academic circles. Himeur et al., (2023) critically analyzed the ethical implications, focusing on data rights, informed consent, and the potential for bias within ML algorithms, highlighting the need for robust ethical standards in industrial data handling [23]. Concurrently, the realm of data security was thoroughly explored by Li et al., (2023), who proposed a comprehensive cybersecurity framework tailored for Big Data environments in production, emphasizing resilience against evolving cyber threats [24].

E. *Integration Hurdles and Scalability Concerns*

The literature is replete with insights into the complexities and challenges facing industries in assimilating these advanced technologies. Stergiou et al., (2023) offered a compelling exploration of the financial and infrastructural impediments that hinder seamless technology adoption, highlighting disparities in readiness levels between large corporations and small-to-medium enterprises (SMEs) [25]. In addition, a survey by Mokhtarimousavi and Mehrabi (2023) provided a global overview of the uneven adoption landscape, suggesting collaborative engagements and policy interventions as vital enablers to bridge this gap [26].

F. *Innovative Approaches and Future Trajectories*

Anticipating future directions, scholars have proposed advanced frameworks and methodologies. An intriguing proposition by Feizizadeh et al. (2023) conceptualized 'adaptive production ecosystems' powered by ML, where production systems autonomously evolve in response to environmental variables, setting the stage for unprecedented operational adaptability [27]. Additionally, Wang et al. (2020) introduced an innovative 'green analytics' model, advocating for sustainable Big Data and ML applications that prioritize energy efficiency and environmental responsibility within production cycles [28].

G. *Theoretical Underpinnings and Conceptual Debates*

Beyond practical applications, the theoretical aspects of integrating Big Data and ML in production have spurred rich academic discussions. Li et al. (2023) contributed significantly to this dialogue, discussing the transformative potential of artificial intelligence and Big Data in production while also cautioning against over-reliance on technology without adequate human oversight [29]. Reinforcing this, a theoretical analysis by Ezugwu et al. (2022) argued for a balanced approach, where technological advancements complement rather than replace human expertise, ensuring sustainable and holistic production ecosystems [30].

H. Theoretical Underpinnings and Conceptual Debates

Empirical studies highlighting real-world applications have significantly enriched the literature. A detailed case study by Mazhar et al. (2023) on the automotive industry showcased how real-time data analytics and ML forecasting models dramatically reduced inventory costs and optimized supply chain operations [31]. Furthermore, a collaborative industry-academic investigation by Bag et al. (2023) into electronics manufacturing illustrated ML's critical role in reducing material waste and improving production line efficiencies through precise demand forecasting and resource allocation [32].

I. Regulatory Frameworks and Compliance Issues

The question of regulatory compliance in the context of Big Data and ML integration has been a focal point in several studies. Regin (2023) explored the legislative landscapes affecting data-driven technologies, highlighting the necessity for dynamic legal frameworks that evolve alongside technological advancements [33]. This perspective was expanded by a compelling study from Goh et al. (2021), which argued for international regulatory harmonization to address the global nature of production networks and the cross-border flow of industrial data [34].

J. Human Factors and Workforce Transformation

Delving into the human aspect, recent studies have illuminated the profound impact of these technologies on the workforce. An insightful analysis by Sharma et al. (2022) presented a dual narrative: while automation may displace certain manual roles, there is a simultaneous creation of new jobs necessitating advanced digital skills, thus urging for proactive workforce retraining initiatives [35]. Complementing this, Fekri et al. (2021) highlighted successful case studies where businesses effectively re-skilled their employees, enabling them to thrive alongside advanced technological integrations [36].

K. Technology Evaluation and Performance Metrics

Scholars have also focused on developing metrics and evaluation protocols for these advanced systems. A notable contribution by Xu et al., (2017) proposed a structured methodology for assessing ML algorithms' performance in production environments, emphasizing accuracy, reliability, and cost-effectiveness [37]. Subsequently, a comprehensive evaluation framework presented by Mazhar et al. (2023) advocated for including adaptability and long-term learning metrics, reflecting the dynamic nature of production settings [38].

L. Stakeholder Engagement and Collaborative Models

The role of diverse stakeholders in steering this technological revolution constitutes a critical narrative within academic contributions. A participatory model proposed by Choi et al. (2022) underscored the necessity for inclusive dialogue, involving policymakers, industry leaders, and academic scholars, to navigate the multifaceted implications effectively and ethically [39]. This model suggests a collective approach to decision-making, ensuring that technological advancements in production management align with broader societal and economic objectives [40].

In conclusion, the extensive body of literature encapsulates the multifaceted nature of Big Data and machine learning integration into production management. It underscores not only the immense potential of these technologies to redefine industrial operations but also the complexities and ethical dimensions requiring careful navigation. Future research endeavors, as suggested by Degrave et al., (2022) and Yu et al., (2021), must continue to unravel these intricate dynamics, drawing upon interdisciplinary insights and fostering collaborative innovation to drive this technological synergy forward sustainably and responsibly [41, 42].

III. Materials and Methods

A. Optimal Production Management

In contemporary industrial contexts, characterized by the prevalence of big data, there has been an expansive diversification in the utilization of data analytics and machine learning across the process industries. This proliferation is visually represented in Fig. 1, delineating the infiltration of these advanced techniques at multiple operational echelons within process-oriented sectors. The scope encompasses both non-interventionist applications manifesting in foundational control loops, such as process surveillance and inferential sensing, and interventionist roles in facets like pinnacle control and strategic decision-making processes [43].

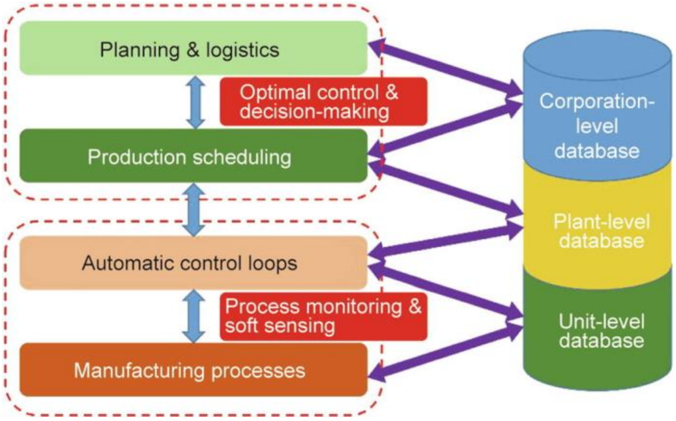


Fig. 1. Optimal production management process.

Non-interventionist applications prioritize providing industry professionals with enhanced perceptual and manipulative command over operational processes. They facilitate the recognition of significant deviations or anomalies, serving a supplementary function without directly initiating process alterations. On the other spectrum, interventionist applications, grounded in data-driven decisions, hold the propensity to command immediate and substantive impacts on the procedural workflow within industrial settings. These decision-making tools, therefore, play a critical role in steering processes, contrasting with their non-interventionist counterparts by directly inducing changes within the industrial operations sphere.

In addressing the enormity and intricacy of medical big data, pharmaceutical entities necessitate specialized analytical mechanisms capable of efficiently navigating and processing this sophisticated data category. Conventional methodologies falter in accommodating the sheer scale of manufacturing data sets, necessitating the exploration of advanced analytical resources, as elucidated in subsequent sections. These big data apparatuses, delineated in Fig. 2, are categorized based on their operational nature into batch processing, real-time (or stream) processing, and interactive analysis. Each category, representing a unique facet of data interaction and manipulation, underscores the multifaceted approach required for the effective assimilation of comprehensive medical data within pharmaceutical research and operational contexts.

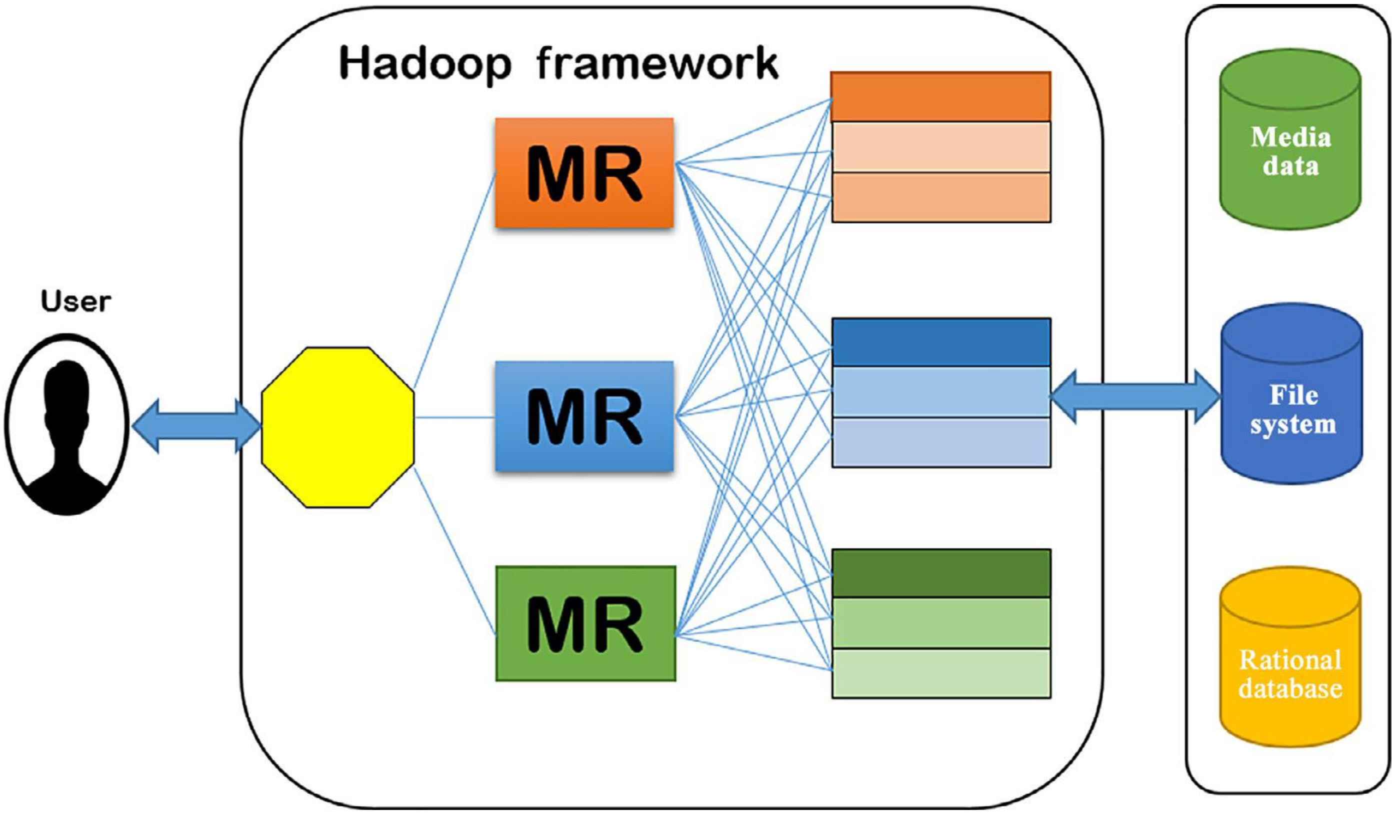


Fig. 2. Apache hadoop software-centric architecture.

Apache Hadoop epitomizes a software-centric architecture, purpose-built to cater to applications demanding extensive data distribution and management. It employs the MapReduce framework, a seminal model delineated in [44], originating from collaborative efforts spearheaded by Google and various contributing entities, to meticulously structure and extrapolate insights from voluminous datasets.

The modus operandi of MapReduce is the strategic decomposition of high-complexity tasks into more manageable fragments. This segmentation process recurs, continually refining the divisions until each constituent issue is sufficiently uncomplicated to be tackled explicitly. Processing clusters are then engaged, operating in a concurrent array to address these distilled sub-issues. This parallel operational structure is pivotal, expediting the computational process by harnessing the collective processing prowess of these clusters. The subsequent phase involves the aggregation of the outcomes produced by these individual processing units, culminating in a synthesized resolution that responds to the initial, more complex query. The intricacies of the Hadoop framework, particularly its function and structural composition, are visually articulated in Fig. 2, providing a detailed schematic of its operational blueprint.

B. *Tools for Optimal Production Management*

1) Big data tools. Big data tools for optimal production management divided into two types as batch processing tools and stream processing tools, and interactive analysis tools as illustrated in Fig. 3. In the prevailing era distinguished by the proliferation of big data, engineers have pioneered the development of open-source frameworks tailored to meet the multifaceted challenges inherent in data-intensive domains. These innovative solutions transcend the realm of batch processing, extending capabilities to encompass stream management and even interactive processing. Such advancements in data interaction techniques empower medical professionals and relevant stakeholders to engage directly with the data repositories. This direct engagement facilitates a more nuanced and individualized analysis, permitting stakeholders to interrogate and interpret the data in alignment with their specific investigative prerequisites. By fostering this level of interaction, these technological enhancements are instrumental in allowing a more refined, requirement-oriented exploration and utilization of extensive data assets within healthcare and related sectors.

2) Stream processing. Stream processing, in the contemporary data paradigm, is integral to managing voluminous data influxes in real-time. Certain applications, spanning industrial sensors, document management, and live online interactions, necessitate the incessant processing of substantial data volumes. Large-scale data, when coupled with the exigencies of real-time processing, mandates minimal latency during its throughput phases [45]. However, the MapReduce framework encounters inherent inefficiencies, particularly a pronounced latency period; data accrued during the 'Map' phase necessitates storage on physical disks prior to initiating the 'Reduce' phase, engendering substantial delays, thus rendering real-time processing impracticable [46].

In the realm of streaming data, the challenges multiply, encompassing issues related to the magnitude of data, accelerated data influx rates, and processing latency. To circumvent the limitations intrinsic to the MapReduce methodology, alternative continuous processing models have gained prominence, such as Storm, Splunk, and Apache Kafka [47]. These innovative platforms are optimized to surmount traditional hurdles by significantly curtailing data transmission delays, thereby facilitating more efficient real-time processing pathways. Consequently, they represent a substantial evolution in tackling the complexities associated with extensive data dimensions, high velocity, and the imperatives of real-time analytics.

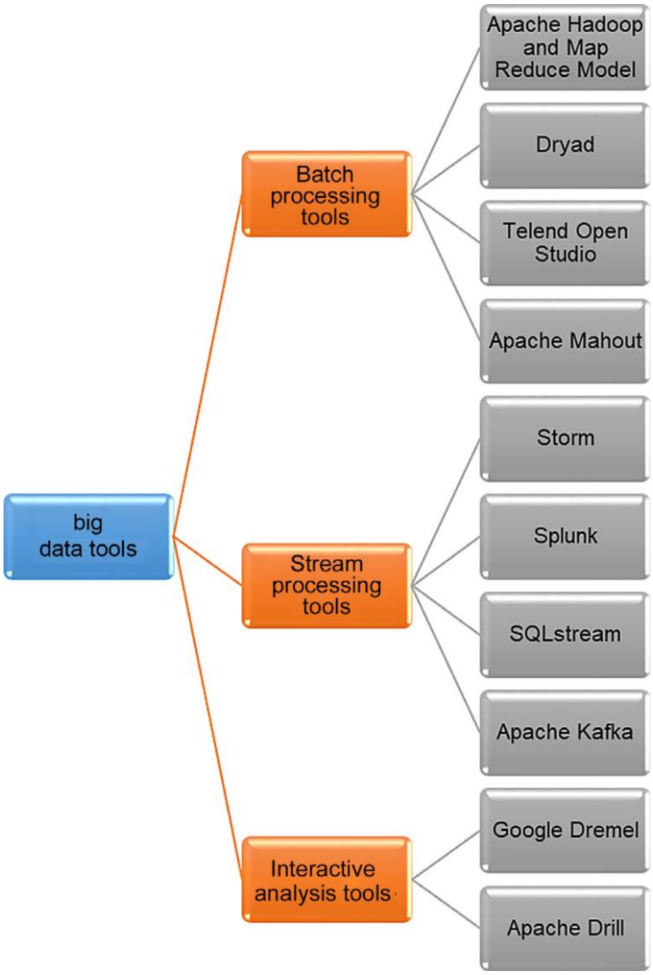


Fig. 3. Big data tools for optimal production management.

3) Interactive analysis tools. In the domain of interactive analysis, especially pertinent to the handling of substantial medical data, the advent of the Apache Drill framework marks a significant evolution. This system, known for its versatility, outstrips counterparts like Google's Dremel, particularly in its capacity to accommodate a variety of query languages, data formats, and sources [48]. Engineered for scalability, Apache Drill is optimized for seamless operation across potentially thousands of servers, proficiently managing data at the byte level, and adeptly handling innumerable user records with minimal latency.

One of the central objectives of Apache Drill is to facilitate the expeditious identification of intersecting data sets, a process crucial for comprehensive data analysis. This functionality distinguishes it within the sphere of large-scale interactive analysis, wherein personalized queries necessitate sophisticated responses, as observed in systems employed by HDFS for storage or intensive batch analysis via the MapReduce framework [49].

Moreover, the prowess of Apache Drill, and similarly advanced platforms like Google's Dremel, lies in their ability to expedite the inquiry process. They enable users to sift through gigabytes of data in response to queries within a matter of seconds, regardless of whether the data is stored in a distributed file system or in a columnar structure. This efficiency underscores the revolution in interactive data analysis, significantly reducing response times and allowing for more nuanced, detailed examinations of colossal data sets.

C. *Applying Deep Learning in Optimal Production Management*

The subsequent sections introduce an innovative framework designed to embed artificial intelligence (AI) methodologies within the Supply Chain Risk Management (SCRM) mechanism, with the primary objective of amplifying the predictive accuracy concerning supply chain threats [50]. This bilateral framework is engineered to foster a collaborative and interactive dynamic between AI specialists and supply chain professionals. In this paradigm, the determinations rendered by AI experts are contingent upon specific, nuanced inputs originating from the supply chain sector. Concurrently, it is imperative that the models devised and the consequent findings generated are sufficiently interpretable to form a solid foundation for, or significantly influence, SCRM decision-making processes.

Fig. 4 delineates the procedural trajectory of the framework. The left segment of the illustration emphasizes the cardinal procedures encapsulated in a data-driven AI approach, while the opposing side outlines the conventional tasks intrinsic to a standard SCRM process. A critical observation is that this framework's architectural integrity hinges on the effective collaboration between two expert cohorts: those versed in data-driven AI techniques and those specializing in supply chain risk management.

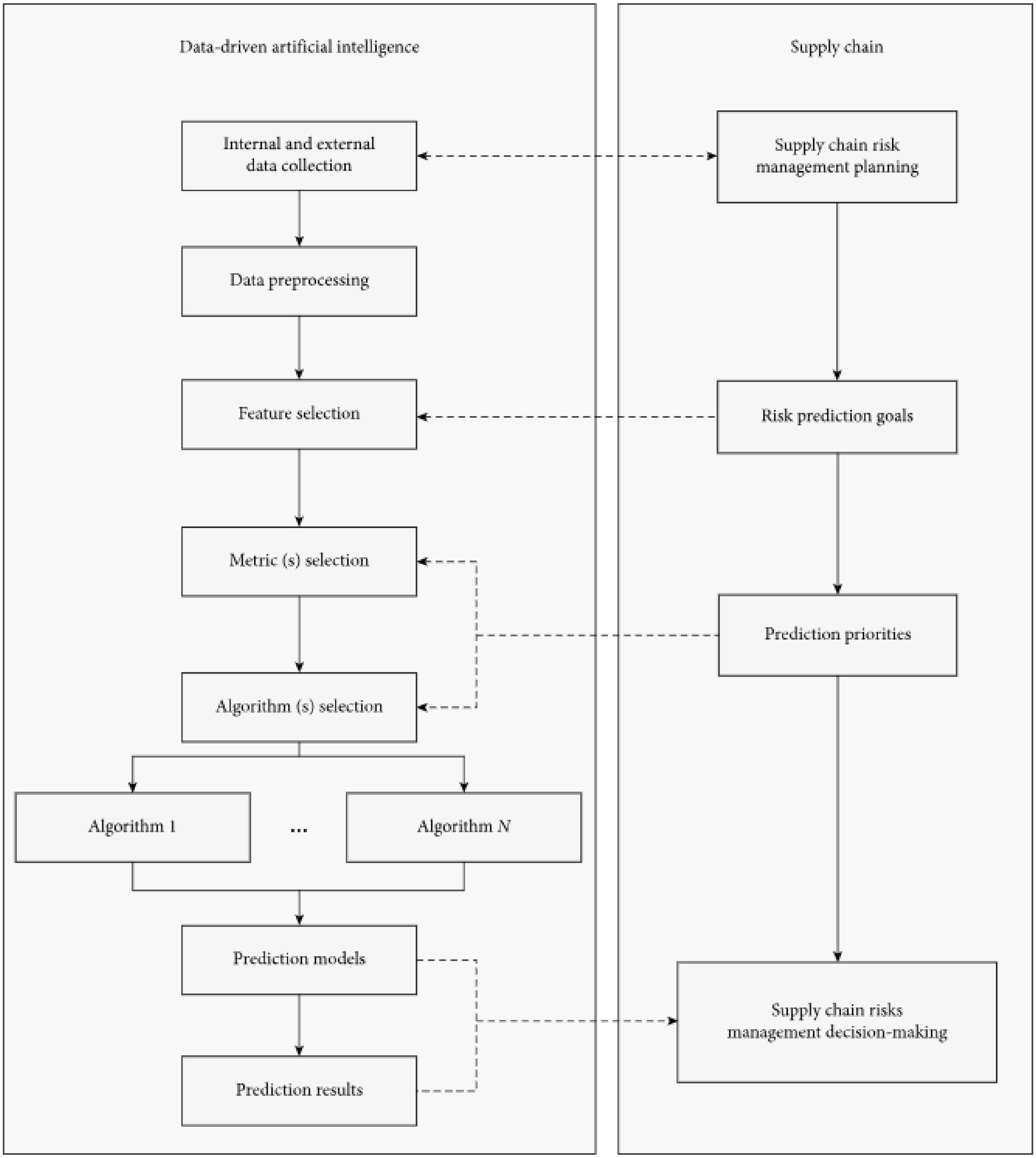


Fig. 4. Data-driven intelligence architecture in big data production management.

By establishing this, the framework ensures a symbiotic relationship wherein both domains leverage their respective expertise, contributing to a more robust, insightful, and responsive risk management strategy. This integrative approach not only enhances the precision of risk forecasting but also fortifies the decision-making apparatus, potentially leading to more secure, efficient, and resilient supply chain networks.

IV. Experimental Results

In the present research, we ventured to integrate advanced big data processing technologies within the context of oil production complications encountered in Kazakhstan. This integration involved the strategic utilization of specified state-of-the-art technologies coupled with innovative techniques meticulously outlined within our study. The primary ambition behind this initiative was the conceptualization and subsequent actualization of a comprehensive framework dedicated to enhancing the management protocols governing oil production activities.

The essence of this framework is captured in Fig. 5, which provides a detailed visual representation of the proposed structural model [51]. This depiction is instrumental in elucidating the functional dynamics and operational interrelationships embedded within the framework, highlighting its potential efficacy in streamlining production management processes.

By harnessing the capabilities of big data, this study underscores a transformative approach in managing the intricacies that characterize the oil production sector in Kazakhstan. The proposed framework, thus, stands as a testament to the potential advancements that could be achieved in production efficiencies, strategic resource allocation, and operational oversight in the oil industry. Moreover, it paves the way for further explorations and potential scalability of similar technologies and methodologies across diverse production landscapes, contributing to a broader narrative of technological integration in industrial practices.

Fig. 6 presents a meticulously organized statistical overview of the proposed framework, articulating complex data in a manner that is both accessible and comprehensible. This deliberate clarity in data visualization is foundational in simplifying the management of voluminous and unstructured information, thereby making the intricacies of big data analytics more approachable.

The utility of Fig. 6 lies in its ability to translate extensive and multifaceted data into intuitive and user-friendly insights [52]. This transformation is crucial for individuals who interact with these datasets, as it demystifies complex patterns and trends within the data, providing stakeholders with a clear vantage point from which to interpret intricate information systems. By distilling this complexity into understandable metrics and visuals, the figure serves as a navigational aid in the decision-making process, enabling stakeholders to make informed decisions grounded in concrete data.

Moreover, the depiction of the framework's statistical data underscores the importance of transparent communication in the realm of big data. It reaffirms the need for tools and methodologies that bridge the gap between complex data management technologies and the individuals who utilize them, ensuring that informed decision-making is not secluded within the realm of data specialists but is a collaborative and inclusive process.

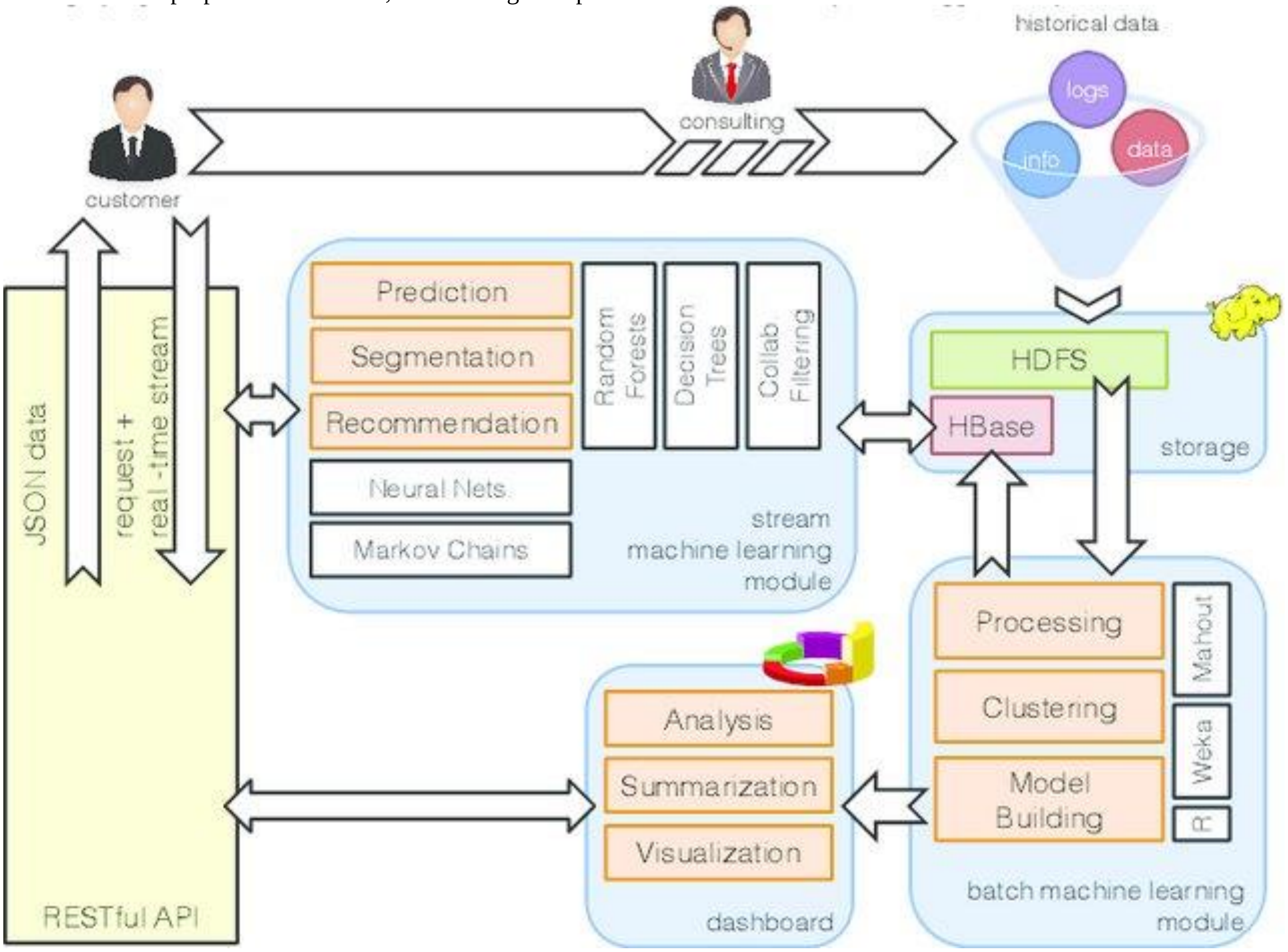


Fig. 5. A framework architecture that supports machine learning based on big data.

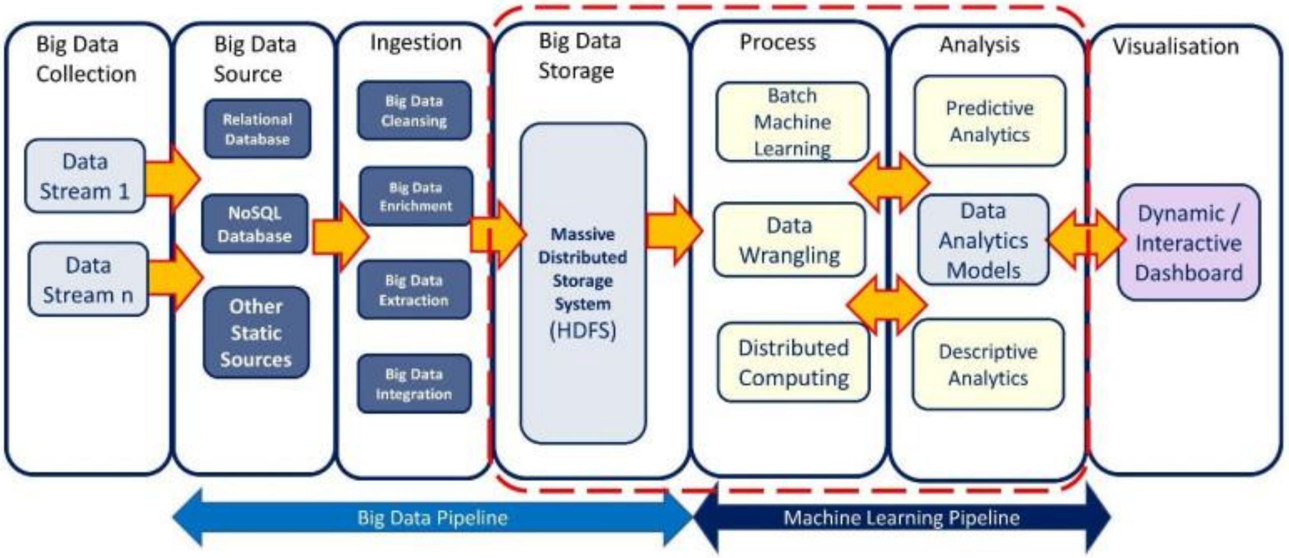


Fig. 6. Facial recognition framework using big data and machine learning.

V. Discussion

The journey through this research has underscored the transformative potential of big data analytics and machine learning (ML) in revolutionizing production management paradigms. By deconstructing the conventional methodologies [53] and introducing a robust framework as shown in earlier sections, the study illuminates the path forward for industries grappling with inefficiencies and the complexities of modern production demands. The nuances of this discussion hinge on the results obtained and their implications, the integration of the framework into existing systems, and the potential challenges and future prospects that industry stakeholders might anticipate.

A. Interpretation and Implications of Results

The results derived from the application of our proposed framework, particularly in the context of oil production in Kazakhstan, as represented in Fig. 5 and Fig. 6, have been nothing short of revelatory. There is an evident enhancement in the management protocols, as seen from the improved statistical information processing and data management capabilities. The framework's ability to process unstructured information efficiently breaks ground in an area where traditional models have consistently stumbled. It harnesses the latent potential within vast data reserves, transforming them into actionable insights that drive strategic decision-making and optimize operational protocols. Furthermore, the dynamics of fuel reserve management, reinforce the framework's utility in planning and resource allocation, critical factors influencing the sustainability and economic feasibility of production endeavors.

The practical implications of these results are manifold. For one, they validate the hypothesis that integrating sophisticated data analysis techniques can tangibly enhance production management. This validation is not merely academic but also carries significant weight for industry stakeholders, potentially influencing policy decisions, investment directions, and strategic business planning. Moreover, the results underscore the need for a paradigm shift in production management, away from traditional, often myopic strategies towards a more integrated, data-driven approach.

B. Integration into Existing Systems

The seamless integration of the proposed framework into existing production management systems is pivotal. This research's applicability hinges on its compatibility with the intricate, multifaceted operational matrices already in place within industries. One of the standout features of the framework is its adaptability, demonstrated through its application in the distinct context of Kazakhstan's oil production industry. However, the integration process poses its own set of challenges, including the need for infrastructural overhaul, upskilling of personnel, and establishment of new oversight and accountability mechanisms.

For industries, the integration also implies a need to re-evaluate and possibly redesign their data infrastructure to accommodate the more sophisticated requirements of big data analytics. This is no small undertaking, as it necessitates both financial investment and a cultural shift towards a more data-centric operational ethos. However, the payoffs, as evidenced by the results, can justify the means, especially in a competitive industrial landscape where efficiency and innovation drive success.

C. Challenges and Limitations

Despite its promising outcomes, the framework's application is not without its challenges. One of the primary constraints is the technological investment required to harness big data fully [54], often a deterrent for smaller enterprises with limited resources [55]. Additionally, while the framework is adaptable, each industry's unique characteristics necessitate a certain degree of customization of the analytics tools and ML algorithms. There is also the human factor to consider, where resistance to change and lack of technical expertise can impede implementation [56].

From a data perspective, issues of privacy, security, and ethical handling of information come to the fore [57-59]. As industries tread the line between data collection for efficiency and violation of privacy norms, a new regulatory landscape may emerge, demanding careful navigation. These challenges are not insurmountable but call for a nuanced understanding and proactive management strategy [60].

D. Future Directions

Looking ahead, the research opens new avenues for exploration. The scalability of the framework across different industry sectors, particularly those not traditionally associated with cutting-edge technology, offers exciting possibilities. Future studies might explore longitudinal impacts, assessing not just immediate productivity gains but also long-term effects on sustainability, employee satisfaction, and consumer responses [61].

Moreover, as technology evolves, so too will the tools at our disposal. Advances in AI, the increasing sophistication of ML algorithms, and improvements in data storage and processing capabilities will continually shape the framework's evolution [62-65]. Further research will need to monitor these trends closely, adapting the framework to remain at the forefront of innovation.

E. Concluding Thoughts

In conclusion, this research marks a significant step forward in our understanding of production management in the age of big data. The proposed framework serves as a beacon, guiding industries towards more efficient, sustainable, and intelligent production methodologies [66-68]. While challenges remain, the potential benefits are undeniable, promising a new era of innovation and excellence in production management [69]. As we stand on the precipice of this new era, the directions we take now will define the industrial landscape of the future.

VI. Conclusion

This research embarked on a journey through the intricate landscape of big data analytics and machine learning, unveiling their profound impact on optimal production management. The study's findings have illuminated the transformative power these technologies hold in redefining traditional methodologies, highlighting an innovative framework adept at harnessing the complexities and vastness of industrial data. The practical trials within the context of Kazakhstan's oil production realm underscored the framework's efficacy, revealing significant enhancements in operational efficiency, resource management, and strategic decision-making. This transition from data to insight represents a critical leap forward, facilitating a more sustainable, responsive, and productive industrial environment.

However, the path ahead is laden with challenges requiring holistic strategies that consider technological, human, and ethical factors. The integration of these advanced systems necessitates not only substantial financial investment but also a paradigm shift in cultural attitudes towards data-driven methodologies. Despite these hurdles, the potential for societal betterment and industrial advancement is palpable, offering a compelling argument for continued exploration and adoption. Future research endeavors in this direction, particularly those focusing on the scalability of the proposed framework and its longitudinal impacts, will be instrumental in steering the evolution of production domains worldwide. As we conclude, it becomes clear that this research is not just an end but a beginning - the inception of a journey toward a new era of industrial revolution propelled by intelligence, efficiency, and foresight.

References

[1] Sheng, J., Amankwah - Amoah, J., Khan, Z., & Wang, X. (2021). COVID - 19 pandemic in the new era of big data analytics: Methodological innovations and future research directions. British Journal of Management, 32(4), 1164-1183.

[2] Shah, H. M., Gardas, B. B., Narwane, V. S., & Mehta, H. S. (2023). The contemporary state of big data analytics and artificial intelligence towards intelligent supply chain risk management: a comprehensive review. Kybernetes, 52(5), 1643-1697.

[3] Rosati, R., Romeo, L., Cecchini, G., Tonetto, F., Viti, P., Mancini, A., & Frontoni, E. (2023). From knowledge-based to big data analytic model: a novel IoT and machine learning based decision support system for predictive maintenance in Industry 4.0. Journal of Intelligent Manufacturing, 34(1), 107-121.

[4] Omarov, B., Altayeva, A., Turganbayeva, A., Abdulkarimova, G., Gusmanova, F., Sarbasova, A., ... & Omarov, N. (2019). Agent based modeling of smart grids in smart cities. In Electronic Governance and Open Society: Challenges in Eurasia: 5th International Conference, EGOSE 2018, St. Petersburg, Russia, November 14-16, 2018, Revised Selected Papers 5 (pp. 3-13). Springer International Publishing.

[5] Ayvaz, S., & Alpay, K. (2021). Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. Expert Systems with Applications, 173, 114598.

[6] Andronie, M., Lăzăroiu, G., Iatagan, M., Uță, C., Ștefănescu, R., & Cocoșatu, M. (2021). Artificial intelligence-based decision-making algorithms, internet of things sensing networks, and deep learning-assisted smart process management in cyber-physical production systems. Electronics, 10(20), 2497.

[7] Tursynova, A., & Omarov, B. (2021, November). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. In 2021 16th International Conference on Electronics Computer and Computation (ICECCO) (pp. 1-4). IEEE.

[8] Kumar, S., Gopi, T., Harikeerthana, N., Gupta, M. K., Gaur, V., Krolczyk, G. M., & Wu, C. (2023). Machine learning techniques in additive manufacturing: a state of the art review on design, processes and production control. Journal of Intelligent Manufacturing, 34(1), 21-55.

[9] Nagy, M., Lăzăroiu, G., & Valaskova, K. (2023). Machine Intelligence and Autonomous Robotic Technologies in the Corporate Context of SMEs: Deep Learning and Virtual Simulation Algorithms, Cyber-Physical Production Networks, and Industry 4.0-Based Manufacturing Systems. Applied Sciences, 13(3), 1681.

[10] Xia, K., Sacco, C., Kirkpatrick, M., Saidy, C., Nguyen, L., Kircaliali, A., & Harik, R. (2021). A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. Journal of Manufacturing Systems, 58, 210-230.

[11] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. Computers, Materials & Continua, 74(3).

[12] Al-Janabi, S., & Al-Janabi, Z. (2023). Development of deep learning method for predicting DC power based on renewable solar energy and multi-parameters function. Neural Computing and Applications, 1-22.

[13] Mitra, A., Bera, B., Das, A. K., Jamal, S. S., & You, I. (2023). Impact on blockchain-based AI/ML-enabled big data analytics for Cognitive Internet of Things environment. Computer Communications, 197, 173-185.

[14] Chen, B., Bai, R., Li, J., Liu, Y., Xue, N., & Ren, J. (2023). A multiobjective single bus corridor scheduling using machine learning-based predictive models. International Journal of Production Research, 61(1), 131-145.

[15] Omarov, B., Suliman, A., & Kushibar, K. (2016). Face recognition using artificial neural networks in parallel architecture. Journal of Theoretical and Applied Information Technology, 91(2), 238

[16] Sahal, R., Breslin, J. G., & Ali, M. I. (2020). Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case. Journal of manufacturing systems, 54, 138-151.

[17] Li, X., Liu, H., Wang, W., Zheng, Y., Lv, H., & Lv, Z. (2022). Big data analysis of the internet of things in the digital twins of smart city based on deep learning. Future Generation Computer Systems, 128, 167-177.

[18] Tseng, M. L., Tran, T. P. T., Ha, H. M., Bui, T. D., & Lim, M. K. (2021). Sustainable industrial and operation engineering trends and challenges Toward Industry 4.0: A data driven analysis. Journal of Industrial and Production Engineering, 38(8), 581-598.

[19] Qi, Q., Xu, Z., & Rani, P. (2023). Big data analytics challenges to implementing the intelligent Industrial Internet of Things (IIoT) systems in sustainable manufacturing operations. Technological Forecasting and Social Change, 190, 122401.

[20] Ming, W., Sun, P., Zhang, Z., Qiu, W., Du, J., Li, X., ... & Guo, X. (2023). A systematic review of machine learning methods applied to fuel cells in performance evaluation, durability prediction, and application monitoring. International Journal of Hydrogen Energy, 48(13), 5197-5228.

[21] Wang, K., Zeng, M., Wang, J., Shang, W., Zhang, Y., Luo, T., & Dowling, A. W. (2023). When physics-informed data analytics outperforms black-box machine learning: A case study in thickness control for additive manufacturing. Digital Chemical Engineering, 6, 100076.

[22] Serey, J., Alfaro, M., Fuertes, G., Vargas, M., Durán, C., Ternero, R., ... & Sabattin, J. (2023). Pattern recognition and deep learning technologies, enablers of industry 4.0, and their role in engineering research. Symmetry, 15(2), 535.

[23] Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., ... & Amira, A. (2023). AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Artificial Intelligence Review, 56(6), 4929-5021.

[24] Li, J., Herdem, M. S., Nathwani, J., & Wen, J. Z. (2023). Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management. Energy and AI, 11, 100208.

[25] Stergiou, K., Ntakolia, C., Varytis, P., Koumoulos, E., Karlsson, P., & Moustakidis, S. (2023). Enhancing property prediction and process optimization in building materials through machine learning: A review. Computational Materials Science, 220, 112031.

[26] Mokhtarimousavi, S., & Mehrabi, A. (2023). Flight delay causality: Machine learning technique in conjunction with random parameter statistical analysis. International Journal of Transportation Science and Technology, 12(1), 230-244.

[27] Feizizadeh, B., Omarzadeh, D., Kazemi Garajeh, M., Lakes, T., & Blaschke, T. (2023). Machine learning data-driven approaches for land use/cover mapping and trend analysis using Google Earth Engine. Journal of Environmental Planning and Management, 66(3), 665-697.

[28] Wang, S., Duan, J., Shi, D., Xu, C., Li, H., Diao, R., & Wang, Z. (2020). A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning. IEEE Transactions on Power Systems, 35(6), 4644-4654.

[29] Li, C., Zheng, P., Yin, Y., Wang, B., & Wang, L. (2023). Deep reinforcement learning in smart manufacturing: A review and prospects. CIRP Journal of Manufacturing Science and Technology, 40, 75-101.

[30] Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence, 110, 104743.

[31] Mazhar, T., Asif, R. N., Malik, M. A., Nadeem, M. A., Haq, I., Iqbal, M., ... & Ashraf, S. (2023). Electric Vehicle Charging System in the Smart Grid Using Different Machine Learning Methods. Sustainability, 15(3), 2603.

[32] Bag, S., Dhamija, P., Luthra, S., & Huisingh, D. (2023). How big data analytics can help manufacturing companies strengthen supply chain resilience in the context of the COVID-19 pandemic. The International Journal of Logistics Management, 34(4), 1141-1164.

[33] Regin, R., Rajest, S. S., & Shynu, T. (2023). A Review of Secure Neural Networks and Big Data Mining Applications in Financial Risk Assessment. Central Asian Journal of Innovations on Tourism Management and Finance, 4(2), 73-90.

[34] Goh, G. D., Sing, S. L., & Yeong, W. Y. (2021). A review on machine learning in 3D printing: applications, potential, and challenges. Artificial Intelligence Review, 54(1), 63-94.

[35] Sharma, P., Said, Z., Kumar, A., Nizetic, S., Pandey, A., Hoang, A. T., ... & Tran, V. D. (2022). Recent advances in machine learning research for nanofluid-based heat transfer in renewable energy system. Energy & Fuels, 36(13), 6626-6658.

[36] Fekri, M. N., Patel, H., Grolinger, K., & Sharma, V. (2021). Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network. Applied Energy, 282, 116177.

[37] Xu, J., Ren, Z., Xie, S., Wang, Y., & Wang, J. (2023). Deep learning - based optimal tracking control of flow front position in an injection molding machine. Optimal Control Applications and Methods, 44(3), 1376-1393.

[38] Mazhar, T., Irfan, H. M., Haq, I., Ullah, I., Ashraf, M., Shloul, T. A., ... & Elkamchouchi, D. H. (2023). Analysis of Challenges and Solutions of IoT in Smart Grids Using AI and Machine Learning Techniques: A Review. Electronics, 12(1), 242.

[39] Choi, T. M., Kumar, S., Yue, X., & Chan, H. L. (2022). Disruptive technologies and operations management in the Industry 4.0 era and beyond. Production and Operations Management, 31(1), 9-31.

[40] Bharadiya, J. P. (2023). A Comparative Study of Business Intelligence and Artificial Intelligence with Big Data Analytics. American Journal of Artificial Intelligence, 7(1), 24.

[41] Degrave, J., Felici, F., Buchli, J., Neunert, M., Tracey, B., Carpanese, F., ... & Riedmiller, M. (2022). Magnetic control of tokamak plasmas through deep reinforcement learning. Nature, 602(7897), 414-419.

[42] Yu, L., Qin, S., Zhang, M., Shen, C., Jiang, T., & Guan, X. (2021). A review of deep reinforcement learning for smart building energy management. IEEE Internet of Things Journal, 8(15), 12046-12063.

[43] Grover, P., Kar, A. K., & Dwivedi, Y. K. (2022). Understanding artificial intelligence adoption in operations management: insights from the review of academic literature and social media discussions. Annals of Operations Research, 308(1-2), 177-213.

[44] Cebekhulu, E., Onumanyi, A. J., & Isaac, S. J. (2022). Performance analysis of machine learning algorithms for energy demand–supply prediction in smart grids. Sustainability, 14(5), 2546.

[45] Benevento, E., Aloini, D., & Squicciarini, N. (2023). Towards a real-time prediction of waiting times in emergency departments: A comparative analysis of machine learning techniques. International Journal of Forecasting, 39(1), 192-208.

[46] Akter, S., Michael, K., Uddin, M. R., McCarthy, G., & Rahman, M. (2022). Transforming business using digital innovations: The application of AI, blockchain, cloud and data analytics. Annals of Operations Research, 1-33.

[47] Ming, W., Sun, P., Zhang, Z., Qiu, W., Du, J., Li, X., ... & Guo, X. (2023). A systematic review of machine learning methods applied to fuel cells in performance evaluation, durability prediction, and application monitoring. International Journal of Hydrogen Energy, 48(13), 5197-5228.

[48] Ageed, Z. S., Zeebaree, S. R., Sadeeq, M. M., Kak, S. F., Yahia, H. S., Mahmood, M. R., & Ibrahim, I. M. (2021). Comprehensive survey of big data mining approaches in cloud systems. Qubahan Academic Journal, 1(2), 29-38.

[49] Taherinezhad, A., & Alinezhad, A. (2023). Nations performance evaluation during SARS-CoV-2 outbreak handling via data envelopment analysis and machine learning methods. International Journal of Systems Science: Operations & Logistics, 10(1), 2022243.

[50] Guo, H. N., Wu, S. B., Tian, Y. J., Zhang, J., & Liu, H. T. (2021). Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. Bioresource technology, 319, 124114.

[51] Baldominos, A., Albacete, E., Saez, Y., & Isasi, P. (2014, December). A scalable machine learning online service for big data real-time analysis. In 2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD) (pp. 1-8). IEEE.

[52] Asaithambi, S. P. R., Venkatraman, S., & Venkatraman, R. (2021). Proposed big data architecture for facial recognition using machine learning. AIMS Electronics and Electrical Engineering, 5(1), 68-92.

[53] Hosseinnia Shavaki, F., & Ebrahimi Ghahnavieh, A. (2023). Applications of deep learning into supply chain management: a systematic literature review and a framework for future research. Artificial Intelligence Review, 56(5), 4447-4489.

[54] Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2023). A review on reinforcement learning algorithms and applications in supply chain management. International Journal of Production Research, 61(20), 7151-7179.

[55] Teh, D., & Rana, T. (2023). The Use of Internet of Things, Big Data Analytics and Artificial Intelligence for Attaining UN's SDGs. In Handbook of big data and analytics in accounting and auditing (pp. 235-253). Singapore: Springer Nature Singapore.

[56] Li, Q., Cui, Z., Cai, Y., Su, Y., & Wang, B. (2023). Renewable-based microgrids’ energy management using smart deep learning techniques: Realistic digital twin case. Solar Energy, 250, 128-138.

[57] Mostafa, N., Ramadan, H. S. M., & Elfarouk, O. (2022). Renewable energy management in smart grids by using big data analytics and machine learning. Machine Learning with Applications, 9, 100363.

[58] Omarov, B., Altayeva, A., Suleimenov, Z., Im Cho, Y., & Omarov, B. (2017, April). Design of fuzzy logic based controller for energy efficient operation in smart buildings. In 2017 First IEEE International Conference on Robotic Computing (IRC) (pp. 346-351). IEEE.

[59] Wang, W., Guo, H., Li, X., Tang, S., Xia, J., & Lv, Z. (2022). Deep learning for assessment of environmental satisfaction using BIM big data in energy efficient building digital twins. Sustainable Energy Technologies and Assessments, 50, 101897.

[60] Wang, J., Xu, C., Zhang, J., & Zhong, R. (2022). Big data analytics for intelligent manufacturing systems: A review. Journal of Manufacturing Systems, 62, 738-752.

[61] Li, C., Chen, Y., & Shang, Y. (2022). A review of industrial big data for decision making in intelligent manufacturing. Engineering Science and Technology, an International Journal, 29, 101021.

[62] Karatas, M., Eriskin, L., Deveci, M., Pamucar, D., & Garg, H. (2022). Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives. Expert Systems with Applications, 200, 116912.

[63] Altayeva, A., Omarov, B., Suleimenov, Z., & Im Cho, Y. (2017, June). Application of multi-agent control systems in energy-efficient intelligent building. In 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS) (pp. 1-5). IEEE.

[64] Ahmad, T., Madonski, R., Zhang, D., Huang, C., & Mujeeb, A. (2022). Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. Renewable and Sustainable Energy Reviews, 160, 112128.

[65] Li, X., Liu, H., Wang, W., Zheng, Y., Lv, H., & Lv, Z. (2022). Big data analysis of the internet of things in the digital twins of smart city based on deep learning. Future Generation Computer Systems, 128, 167-177.

[66] Kliestik, T., Zvarikova, K., & Lăzăroiu, G. (2022). Data-driven machine learning and neural network algorithms in the retailing environment: Consumer engagement, experience, and purchase behaviors. Economics, Management and Financial Markets, 17(1), 57-69.

[67] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019). Applying face recognition in video surveillance security systems. In Software Technology: Methods and Tools: 51st International Conference, TOOLS 2019, Innopolis, Russia, October 15–17, 2019, Proceedings 51 (pp. 271-280). Springer International Publishing.

[68] Zhou, L., Jiang, Z., Geng, N., Niu, Y., Cui, F., Liu, K., & Qi, N. (2022). Production and operations management for intelligent manufacturing: A systematic literature review. International Journal of Production Research, 60(2), 808-846.

[69] Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., ... & Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. Automation in Construction, 141, 104440.

Development of an Intelligent Service Delivery System to Increase Efficiency of Software Defined Networks

Serik Joldasbayev[1], Saya Sapakova[2], Almash Zhaksylyk[3],
Bakhytzhan Kulambayev[4], Reanta Armankyzy[5], Aruzhan Bolysbek[6]
International Information Technology University, Almaty, Kazakhstan[1]
Al-Farabi Kazakh National University, Almaty, Kazakhstan[1]
Department of Computer Engineering, International University of Information Technology, Almaty, Kazakhstan[2, 5]
Satbayev University, Almaty, Kazakhstan[3]
Turan University, Almaty, Kazakhstan[4]
Bachelor Student at International Information Technology University, Almaty, Kazakhstan[6]

Abstract*—The burgeoning complexity in network management has garnered considerable attention, specifically focusing on Software-Defined Networking (SDN), a transformative technology that addresses limitations inherent in traditional network infrastructures. Despite its advantages, SDN is often susceptible to bottlenecks and excessive load issues, underscoring the necessity for more robust load balancing solutions. Previous research in this realm has predominantly concentrated on employing static or dynamic methodologies, encapsulating only a handful of parameters for traffic management, thereby limiting their effectiveness. This study introduces an innovative, intelligence-led approach to service delivery systems in SDN, specifically by orchestrating packet forwarding—encompassing both TCP and UDP traffic—through a multi-faceted analysis utilizing twelve distinct parameters elaborated in subsequent sections. This research leverages advanced machine learning algorithms, notably K-Means and DBSCAN clustering, to discern patterns and optimize traffic distribution, ensuring a more nuanced, responsive load balancing mechanism. A salient feature of this methodology involves determining the ideal number of operational clusters to enhance efficiency systematically. The proposed system underwent rigorous testing with an escalating scale of network packets, encompassing counts of 5,000 to an extensive 10,000,000, to validate performance under varying load conditions. Comparative analysis between K-Means and DBSCAN's results reveals critical insights into their operational efficacy, corroborated by juxtaposition with extant scholarly perspectives. This investigation's findings significantly contribute to the discourse on adaptive network solutions, demonstrating that an intelligent, parameter-rich approach can substantively mitigate load-related challenges, thereby revolutionizing service delivery paradigms within Software-Defined Networks.

***Keywords*—Load balancing; machine learning; server; classification; software**

I. INTRODUCTION

In the contemporary digital landscape, the explosive growth in data traffic, coupled with the increasing reliance on cloud services and the Internet of Things (IoT), has rendered traditional networking architectures both obsolete and inadequate [1]. These legacy systems, characterized by their rigidity and hardware-dependency, present significant hurdles in catering to the dynamic nature of modern data traffic and the need for real-time decision-making [2]. It is within this challenging context that Software-Defined Networking (SDN) [3] has emerged as a beacon of innovation, enabling unprecedented levels of network management and adaptability by abstracting the control logic from the underlying physical infrastructure.

However, the benefits of SDN, particularly its centralized control and programmable network behavior, also bring forth complex challenges, chief among them being effective load balancing [4]. The conventional load balancing mechanisms, with their static nature, fail to comprehend and adapt to the erratic behavior of contemporary network traffic [5], leading to suboptimal utilization of network resources, potential network bottlenecks, and compromised service quality [6]. Thus, there is an exigent need for a more intelligent, scalable, and adaptive load balancing solution, capable of interpreting complex network environments and autonomously optimizing traffic distribution to enhance overall network performance.

Recognizing these challenges, this research paper explores the integration of machine learning (ML) techniques into the SDN architecture, specifically aimed at revolutionizing the load balancing processes [7]. Machine learning, with its capability to analyze vast datasets [8], identify patterns [9], and make predictive decisions [10], presents a promising solution to the intricate problem of dynamic load balancing. By applying ML algorithms, the research intends to enable SDN controllers to make real-time traffic routing decisions based on data-driven insights, thereby significantly improving resource allocation [11], reducing latency [12], and enhancing the user experience [13].

The paper begins by establishing the foundational concepts critical to this discourse, including an overview of Software-Defined Networking (its architecture, functionalities, and significance) and the inherent challenges of traditional load balancing strategies. This backdrop is essential for understanding the transformation that SDN brings to network management and the subsequent complexities, particularly in maintaining efficient traffic flow and equitable server workloads.

Following this, the discussion shifts focus to the core proposition of this study: the application of machine learning in enhancing SDN load balancing. Herein, the paper will dissect various machine learning models suitable for this application, considering their strengths and potential limitations in real-time decision-making and predictive analysis. The discussion extends to the adaptation of these ML models within the SDN framework, detailing the process from the initial stages of data collection and processing, to the advanced stages of algorithm training, validation, and deployment.

In application, the integration of ML into SDN for load balancing manifests as a dynamic, self-learning system capable of monitoring network conditions, predicting traffic fluctuations, and preemptively redistributing loads across servers to prevent imbalance and congestion [14]. The system's ability to learn continuously from network behavior and traffic patterns marks a significant advancement over traditional methods, essentially transforming reactive responses into proactive strategies.

To substantiate the theoretical discourse, the paper will present empirical evidence derived from simulated SDN environments, demonstrating the practical efficacy and reliability of ML-augmented load balancing [15]. These experiments highlight the performance improvements in terms of reduced network latency, efficient resource utilization, enhanced traffic management, and overall service quality.

Conclusively, this research contributes to the academic and practical realms of network management by advocating for a synergistic approach between machine learning and Software-Defined Networking. The insights drawn from this study underscore the potential of machine learning not just as a tool for load balancing, but as a comprehensive solution for creating intelligent, self-sustaining, and highly efficient network infrastructures, setting a new standard for future network operations and research endeavors.

II. Related Works

A. Understanding Software-Defined Networking and the Load Balancing Dilemma in Servers

The transformation of networking through Software-Defined Networking (SDN) is well-documented in literature, providing a shift from traditional, hardware-centric networks towards a flexible, software-driven approach [16]. The SDN model, advocating for the separation of the control plane from the data plane, introduces programmability into network management, thereby offering centralized control and real-time decision-making [17]. Despite its advanced capabilities, SDN faces inherent challenges, particularly in load balancing, a critical component for maintaining efficient server performance and network stability [18]. Fig. 1 demonstrates traditional IP network load balancing architecture.

In the realm of server environments, especially, load balancing acts as a linchpin, determining the operational robustness and reliability of network services. Traditional load balancing methods, as critiqued in [19], often rely on predetermined, static policies, lacking the flexibility and intelligence to adapt to changeable network traffic, leading to issues like server overload or underutilization. Moreover, the study in [20] highlights the inefficacy of these methods in contemporary data-intensive scenarios, underscoring the need for more sophisticated, context-aware solutions. Fig. 2 demonstrates a software defined network model for load balancing.

B. Machine Learning – A Paradigm Shift in Network Management

Exploring beyond conventional methodologies, recent studies have illuminated the role of machine learning (ML) in redefining network management. The comprehensive review in [21] illustrates ML's capabilities in pattern recognition, anomaly detection, and predictive analysis, marking a significant departure from rule-based systems towards adaptive, autonomous operations. Specifically, machine learning algorithms can analyze extensive datasets, learning and evolving through experiences without explicit programming for every contingency [22].

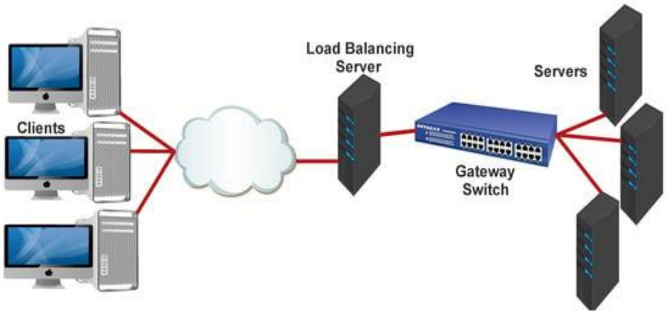


Fig. 1. Topology of self-organizing map.

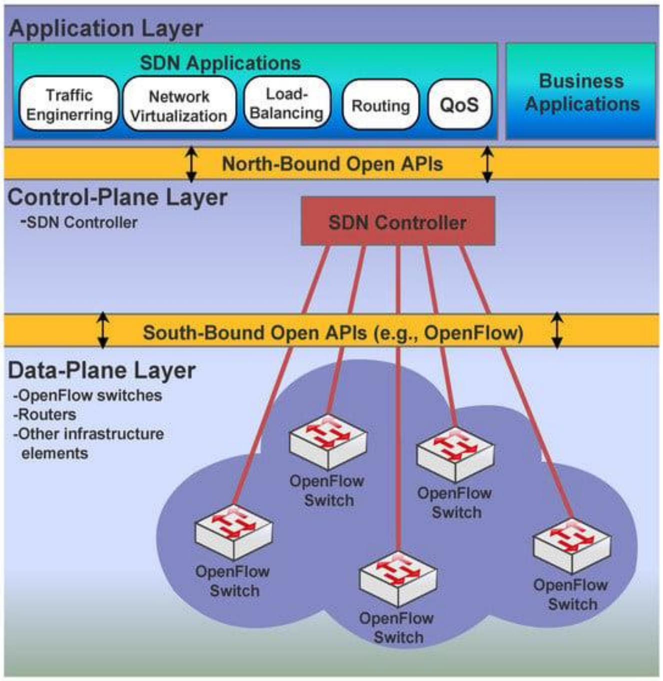


Fig. 2. Software defined network model for load balancing.

In study [23], the authors evaluate various machine learning models, highlighting their suitability for different network scenarios based on accuracy, computational requirements, and ease of implementation. These models, ranging from supervised learning algorithms like decision trees and neural networks to unsupervised techniques like clustering, have found applications across network security, traffic classification, and resource management [24]. Particularly, the predictive and adaptive nature of ML models positions them as ideal candidates for managing unpredictable, high-volume network traffic in real-time [25].

C. Synergizing Machine Learning with SDN for Enhanced Load Balancing

The intersection of machine learning with SDN, especially in the context of load balancing, is a relatively nascent yet rapidly evolving field of research. Several pioneering studies have demonstrated the feasibility and benefits of this integration. For instance, the work in [26] introduces an ML-based framework that empowers the SDN controller with data-driven intelligence to dynamically manage network loads, optimizing both the distribution and routing of traffic. By monitoring network conditions and making informed decisions, this approach mitigates common issues such as bottlenecks and uneven server workloads.

A similar study in [27] exploits the predictive capabilities of ML to forecast traffic patterns and potential hotspots in the network, enabling proactive load balancing measures before servers are critically impacted. Here, machine learning models are trained using historical network data, achieving the foresight necessary for anticipating and preparing for future traffic conditions. Another notable research [28] adopts reinforcement learning, an approach where algorithms learn optimal strategies through trial and error, fostering a self-adjusting, resilient load balancing mechanism. Fig. 3 demonstrates an intelligent software defined network architecture with machine learning techniques.

D. Potential Challenges and Ethical Implications

However, the implementation of ML-based solutions in SDN load balancing is not without its challenges. References [29] and [30] discuss technical hurdles, including the need for large training datasets, algorithmic complexity, and the intensive computational power required for real-time analysis and decision-making. Furthermore, concerns regarding data privacy and security emerge, considering the sensitive nature of network traffic data, necessitating robust encryption and privacy preservation techniques [31].

The literature also addresses the broader ethical implications of integrating ML into network systems. The autonomous decision-making aspect, while contributing to efficiency, raises questions about accountability and transparency in machine decisions, especially in scenarios leading to service denial or prioritization [32]. Studies like [33] argue for the establishment of ethical guidelines and regulatory frameworks to ensure that ML-driven networking adheres to principles of fairness, privacy, and non-discrimination.

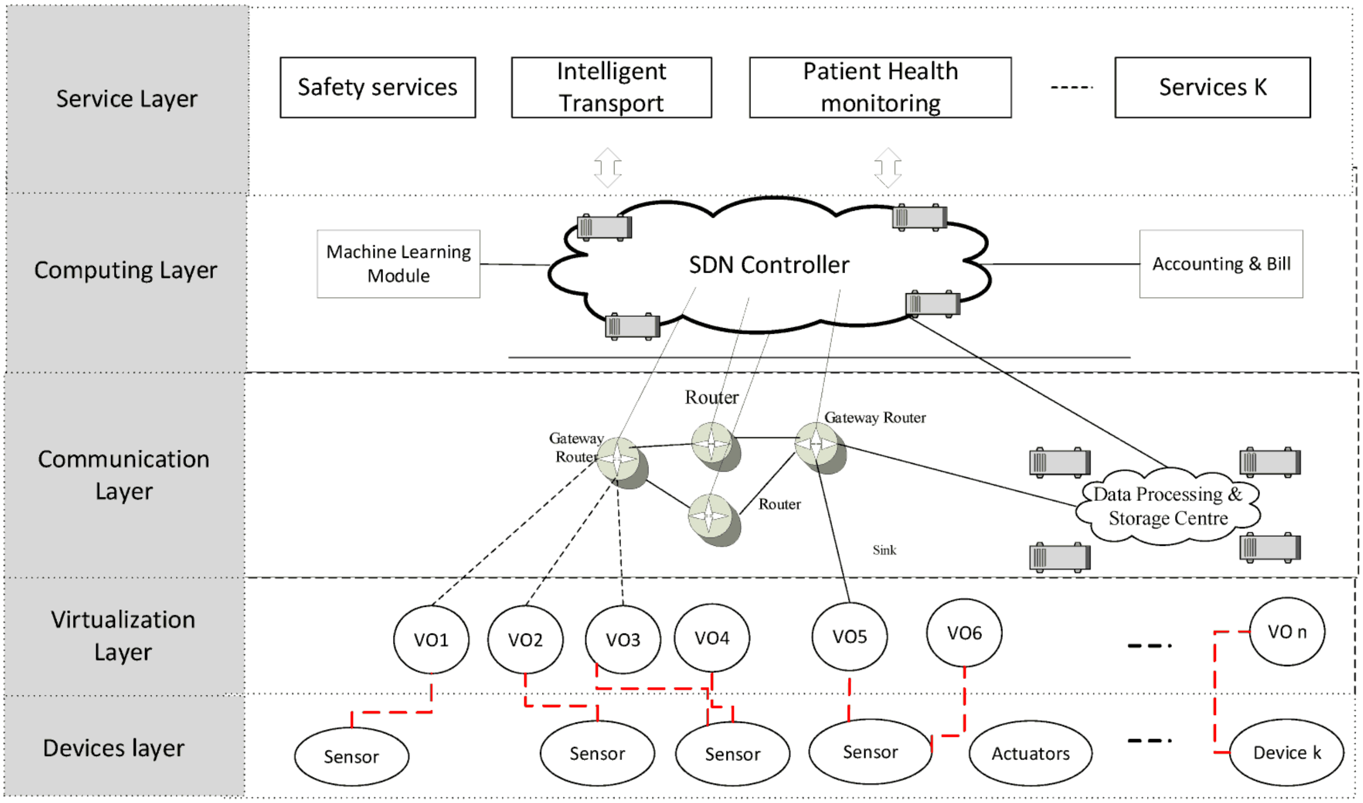


Fig. 3. Intelligent software defined network architecture with machine learning.

In conclusion, the synergy between machine learning and Software-Defined Networking opens a new frontier in network management, particularly in addressing the perennial issue of load balancing. While current research, as explored, has laid a substantial groundwork in this field, indicating notable improvements in network performance and resource optimization, it is evident that further explorations and solutions are necessary [34-36]. Future studies will need to delve deeper into overcoming the practical and ethical challenges present, pushing the boundaries to realize the full potential of ML-integrated SDN systems. This endeavor not only involves the technical refinement of algorithms and systems but also the establishment of standards and protocols that align with ethical norms and societal values.

III. MATERIALS AND METHODS

In this study, we employed Jupyter Notebook as the primary platform for executing our research methodology, aiming to innovate within the realm of Software-Defined Networking (SDN) by integrating a multifaceted parameters approach. The data pivotal to our research was sourced from Kaggle.com, originating from Universidad Del Cauca Popayan, Colombia, which provided a rich foundation for our empirical analysis. A visual representation of the methodology adopted is detailed in Fig. 4, elucidating the sequential steps involved in the proposed technique.

The essence of our proposed methodology initiates when a client propels a request targeted at a specific service from the server. This request is strategically directed first to the Software Load Balancer, where our uniquely designed algorithm is embedded. The initial phase of the algorithmic process involves the extraction of flow statistics, encompassing elements such as IP addresses, port data, inter-arrival times, and more, a task accomplished utilizing the CIC Flowmeter [30].

Subsequent to the accumulation of these integral features, the algorithm proceeds to compute the cluster value associated with the incoming request. This calculated value plays a crucial role in the ensuing step, where the request is systematically channeled to the most fitting server, optimizing efficiency and resource allocation [31]. This mechanism ensures not only equitable server load distribution but also a refined, responsive interaction between the client and server systems. Fig. 5 illustrates of topology of self-organizing map.

The forthcoming sections of this paper promise a comprehensive dissection of each component within our methodology. By unfolding the intricate layers of our approach, we intend to provide clear insights into the functionality of the proposed system, emphasizing its potential to revolutionize load balancing in Software-Defined Networks through intelligent, parameter-sensitive mechanisms [32]. This exploration will underline the practical implications of our research, contributing significantly to the existing body of knowledge in this dynamic field of study.

A. Proposed Method

Within the framework of our proposed methodology, a distinct approach is employed concerning the handling of client-initiated requests for specific services directed at the server. Initially, these requests are routed towards the Software Load Balancer, which incorporates a specially designed algorithm for this precise function. The primary action of this algorithm involves the acquisition of detailed flow statistics through the application of the CIC Flowmeter. This step is critical as it gathers comprehensive data, including various nuanced network traffic features essential for the subsequent stages. Fig. 6 demonstrates an algorithmic structure of the proposed method.

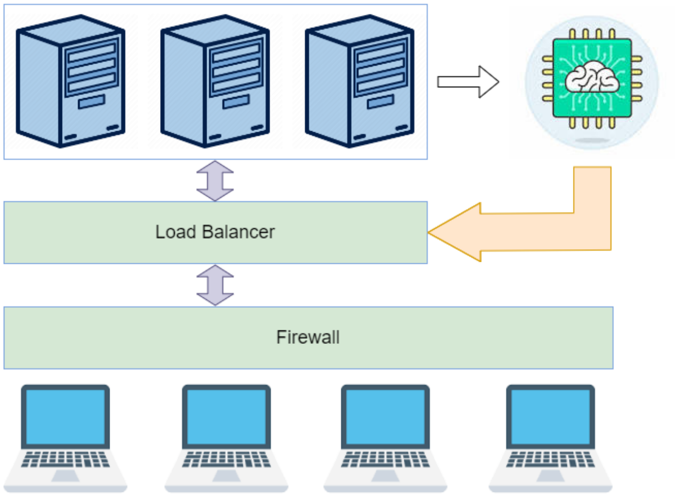


Fig. 4. Sequential steps of a load balancer.

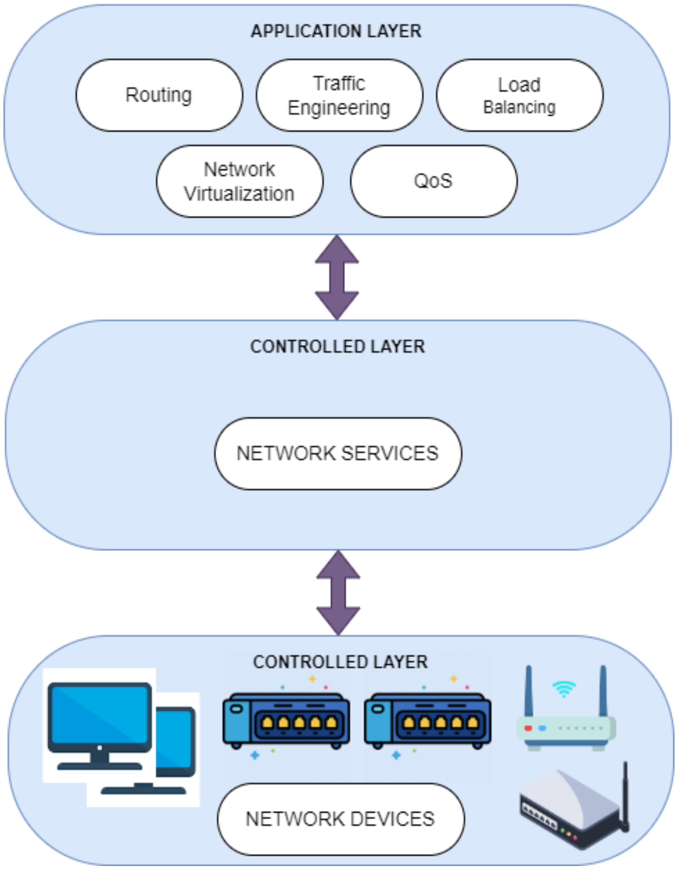


Fig. 5. Topology of self-organizing map.

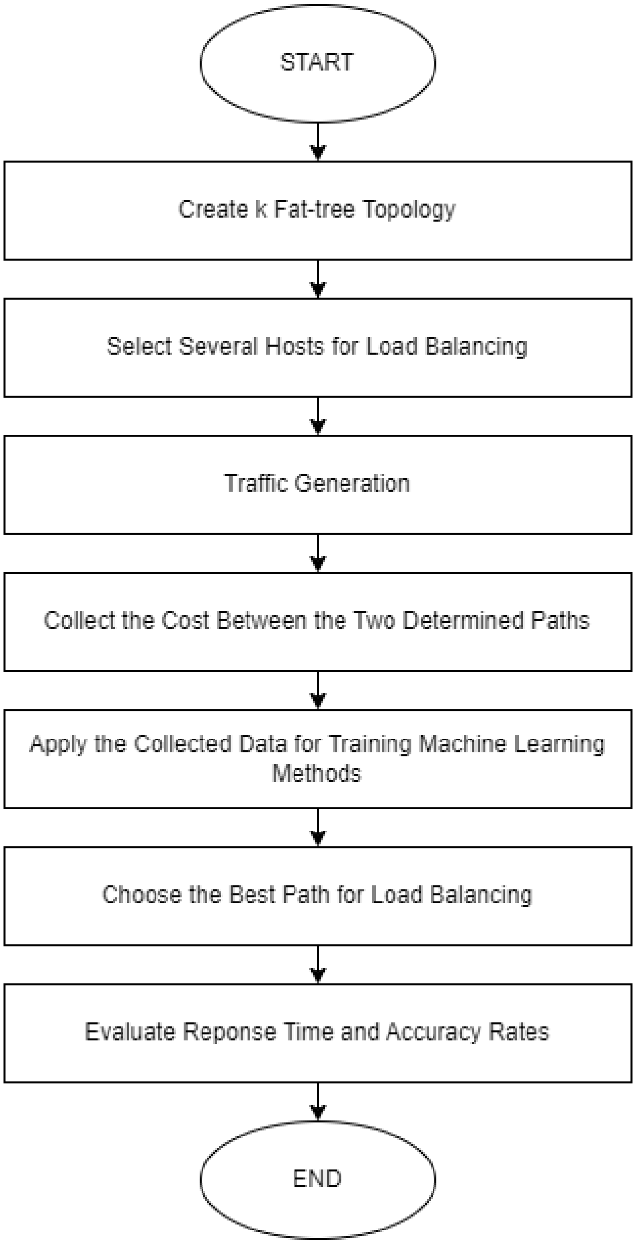

START
Create k Fat-tree Topology
Select Several Hosts for Load Balancing
Traffic Generation
Collect the Cost Between the Two Determined Paths
Apply the Collected Data for Training Machine Learning Methods
Choose the Best Path for Load Balancing
Evaluate Reponse Time and Accuracy Rates
END

Fig. 6. Algorithmic structure of the proposed method.

Following the data collection, a sophisticated process is initiated where the algorithm, utilizing techniques from the KMeans clustering method, diligently computes the cluster value corresponding to each request [33]. This phase is integral for the intelligent distribution of network requests. In conjunction, the system conscientiously assesses the current load of the targeted server, specifically gauging the number of requests it is currently processing against its known capacity thresholds.

If the server's load is ascertained to be below the predefined acceptable threshold, the request is then responsibly forwarded [34]. However, if the contrary is identified — indicating a potential overload scenario — the system activates a contingency mechanism. Instead of overburdening the server, the request is strategically redirected to intermediate nodes within the network infrastructure [35]. This judicious decision ensures the prevention of any single point's overload, thereby maintaining a harmonious, efficient network flow and service delivery.

This intricate process exemplifies the core operational strategy of our load balancing methodology, tailored for Software-Defined Networks (SDN) and enhanced via machine learning techniques. The subsequent section of this study will delve deeper into the specifics of the dataset employed, illuminating how this foundational element plays a pivotal role in informing and guiding the algorithm's decision-making processes. Through this, we aim to underscore the feasibility and effectiveness of integrating advanced machine learning techniques into traditional load balancing practices, setting a new benchmark for efficiency and intelligent network management.

B. *Dataset*

In the course of our research, we sourced our data from an extensive dataset available on the Kaggle platform, originating from network traffic captured at Universidad Del Cauca in Popayan, Colombia, specifically during the morning and evening hours of the year 2017 [36-37]. This comprehensive dataset encompasses a total of 3,577,296 captured packets. For the purposes of our study, a subset consisting of 100,000 packets was extracted to serve as the training material for our machine learning model.

The original dataset is characterized by a complexity of 87 distinct features, each providing insights into various aspects of the network packets. However, for the scope of our work focused on load balancing, we distilled our feature set, selecting 12 critical features instrumental in our analysis and subsequent processing [38]. These features, detailed henceforth, provide the granular data necessary for the intricate understanding and effective management of network load dynamics:

Source.IP: Represents the IPV4 address originating from the client.

Destination.IP: Specifies the IPV4 address intended for receipt by the destination.

Source.Port: Denotes the unique port number used by the source.

Destination.Port: Identifies the respective port number at the destination.

Flow.Duration: Captures the cumulative duration of the flow in milliseconds (ms).

Flow.IAT.Std: Measures the standard deviation of packet inter-arrival times within the flow, reflecting the variability in packet transmission intervals.

Fwd.IAT.Std: Indicates the standard deviation of inter-arrival times of packets forwarded from source to destination, offering insights into the consistency of sent packets.

Bwd.IAT.Std: Contrasts the above by presenting the standard deviation of inter-arrival times for packets received from the destination.

Packet.Length.Std: Accounts for the standard deviation of packet lengths, acknowledging the variability in the size of packets traversing the network.

Down.Up.Ratio: Analyzes the ratio of download to upload traffic, a critical factor in understanding network load.

Active.Std: Observes the standard deviation of the time periods wherein the packets are actively engaged in transmission before transitioning to inactivity.

Idle.Std: Mirrors the active standard deviation by monitoring the idle periods preceding packet transmission activity.

This illustration is imperative, granting stakeholders an overview of the data's distribution and tendencies, thereby underlining the empirical foundation upon which subsequent load balancing strategies are developed and validated. This rigorous approach ensures that the machine learning model is not only grounded in reality but also attuned to the nuances of network traffic behavior, essential for the optimization of load balancing in Software-Defined Networks [39].

In this study, an intricate understanding of the dataset's characteristics is pivotal, ensuring a robust foundation for subsequent analyses. This is achieved through a detailed statistical breakdown of the dataset, focusing on several key metrics that offer insights into each data column's properties [40]. These specific parameters are integral to comprehending the underlying data structure and are crucial for the preparatory phases of data handling, particularly data preprocessing [41-45]. The parameters include:

Count: This represents the total number of entries or elements present in a specific column. It is fundamental in identifying the data volume and assessing if any missing values are in the dataset that may skew analysis or require imputation.

Mean: This metric provides the average value of the data in a particular column, offering a central value around which the data points tend to cluster. This is crucial for understanding the typical or 'normal' value, helping identify outliers or trends that deviate from the expected pattern.

Standard Deviation (std): The standard deviation indicates the dispersion or variability of the data points in a column around the mean. A higher standard deviation signifies greater variability, and conversely, a lower standard deviation suggests that the values cluster closely around the mean. Understanding this aspect is vital for predicting the consistency of data and managing expectations for anomaly detection.

Minimum (min): This value highlights the smallest number recorded in a data column. Identifying minimum values is critical, especially in scenarios where certain thresholds or limits must not be breached for operational or security reasons.

Percentiles (25th, 50th, and 75th): These values delineate the distribution of data in quartiles, indicating where a particular data point stands relative to the rest of the dataset. The 25th percentile represents the lower quartile, the 50th percentile (or median) denotes the middle value, and the 75th percentile signifies the upper quartile. Analyzing these metrics assists in understanding the data's spread, skewness, and the presence of potential outliers.

Maximum (max): Conversely to the minimum, this value represents the highest value in a column. Recognizing the maximum values is essential, particularly in contexts where operations or functionalities are sensitive to value surges, requiring optimizations or safeguarding measures.

These comprehensive statistics serve as the bedrock for data preprocessing, as they collectively offer a multi-dimensional view of the data's behavior. By understanding these fundamental aspects, researchers and analysts can make informed decisions in the subsequent stages of the study. The ensuing section delves deeper into 'Data Preprocessing,' where these statistical insights are employed to refine the dataset, ensuring it is primed for the ensuing stages of machine learning model development and deployment. This process is critical, as the quality of data preprocessing significantly influences the accuracy and reliability of outcomes in data-driven initiatives.

C. Data Preparation

Data preprocessing stands as a cornerstone in the realm of data analytics and machine learning, pivotal in refining raw data into a more digestible and analyzable format, enhancing the efficacy of subsequent operations [46]. This transformative process encompasses several stages, each critical to enhancing the data's quality and, consequently, the outcomes derived from it [47-49]. The stages integral to data preprocessing include:

Data Cleaning: This initial stage addresses the dataset's hygiene, identifying and rectifying missing, incomplete, or irrelevant records and anomalies. It ensures coherence and consistency in the dataset, preparing it for more accurate analysis. Methods like imputation, noise filtering, or outright removal of corrupted data fall under this category.

Data Transformation: Post-cleaning, data transformation or scaling procedures are employed, often to streamline the data attributes across a common scale or format. This harmonization is crucial for algorithms to interpret the data accurately, ensuring that variances in measurements, magnitudes, or formats don't lead to biases or misinterpretations. Techniques such as normalization, aggregation, or encoding categorical variables are typical examples of data transformation.

Data Reduction: Given the extensive volumes of data in contemporary use, data reduction techniques are essential to distill the information and retain only the most relevant attributes for analysis. This step enhances computational efficiency and focuses the analysis on the data aspects most critical to the objectives. Methods like dimensionality reduction, feature selection, and binning are commonly utilized for this purpose.

In the context of our research methodology, these preprocessing steps are meticulously applied. Initially, an examination for missing or noisy data is conducted, alongside an assessment of each parameter's data type. This evaluation is facilitated by modules from Python's extensive libraries, particularly pandas for data manipulation and Seaborn for statistical visualization.

Upon identifying irregularities, we implement hashing on the datasets using the 'apply map()' method inherent in pandas, converting data into a fixed-size value or index. This process streamlines the data structure, enhancing its manageability and security.

Subsequently, to ensure uniformity in data scales, especially when dealing with parameters with varying measurement units or ranges, normalization is carried out. Here, we employ the min-max scaling technique, a standard form of feature scaling that restructures the data values so they reside within a specified range, typically between 0 and 1. This scaling is vital, making sure no particular feature dominates others due to its scale, and it works by re-scaling the range of features while maintaining the same relative distribution and relation between values.

The min-max scaler operates on a straightforward principle, adjusting values along the dataset's minimum and maximum values according to the formula [50]:

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where, Xnorm is the normalized value, X represents the original value, and Xmin and Xmax denote the minimum and maximum values in the original data feature, respectively.

This formula's application rescales the data, ensuring a balanced representation across features, enhancing the subsequent machine learning models' performance by reducing the potential bias that disproportionate data might introduce. The ensuing sections will delve into how these preprocessed data contribute to building a robust, intelligent system.

D. Machine Learning Model Training

In the realm of machine learning, particularly in clustering methodologies, understanding the concept of distance is crucial as it is fundamentally linked to how algorithms, such as KMeans and DBSCAN, group data points into distinct clusters based on inherent similarities or differences [51]. Both these algorithms employ the concept of Euclidean distance, a measure indicative of the 'straight line' distance between two points in a multi-dimensional space, often visualized through geometric planes in basic examples, but in practice, applicable to spaces with any number of dimensions.

Euclidean distance is pivotal because it quantifies the disparity between two entities, which is foundational for clustering. Essentially, a smaller Euclidean distance between two points signifies greater similarity, indicating that they belong to the same cluster or group. Conversely, a larger Euclidean distance is indicative of considerable dissimilarity, often signifying that the points pertain to different clusters [52].

In an n-dimensional space, which is common in machine learning applications due to the multiple features or variables considered, the Euclidean distance between two points—say, Point A and Point B—is calculated using the following mathematical formula [53]:

$$d = \sqrt{\left(A_1 - B_1\right)^2 + \left(A_2 - B_2\right)^2 + \ldots + \left(A_n - B_n\right)^2} \tag{2}$$

Or more succinctly,

$$d = \sqrt{\sum_{i=1}^{n} \left(A_i - B_i\right)^2} \tag{3}$$

Here, d represents the Euclidean distance, while Ai and Bi correspond to the coordinates of Points A and B, respectively, in the n-dimensional space. Each term within the summation accounts for a single dimension's contribution to the total distance, and the aggregation of these individual disparities across all dimensions provides a comprehensive measure of the overall distance between the two points.

This calculation becomes particularly significant in our context, where KMeans and DBSCAN are utilized for clustering based on flow statistics represented through 12 parameters. By leveraging Euclidean distance, these algorithms can effectively assess and quantify the similarity or dissimilarity between various data points (or flows), subsequently enabling the informed and logical grouping of these points into cohesive clusters.

This clustering, based on quantifiable measures, ensures a methodical and data-driven approach to categorization, essential for nuanced tasks like network traffic management, anomaly detection, or optimization in a Software-Defined Network (SDN) setting. It forms the basis for further analyses and decision-making processes that hinge on the understanding of data relationships and classifications in the multidimensional space that the dataset occupies.

IV. Experimental Results

In the investigative analysis encompassing the study, distinct scenarios were methodically examined to elucidate the performance metrics of various advanced computational algorithms, specifically focusing on Logistic Regression (LR), Support Vector Machine (SVM) with hyperparameters C valued at 1 and 100 respectively, Artificial Neural Networks (ANN), and Deep Learning (DL) techniques [54-56]. These algorithms were subjected to rigorous evaluation to discern their efficacy and responsiveness within the constructed scenarios.

The first scenario's results are meticulously documented, with the response times explicitly recorded in milliseconds. These outcomes are visually represented in Fig. 7. This format provides a clear, comparative analysis of the algorithms' performance under the conditions stipulated in the first scenario.

Conversely, the second scenario, crafted to perhaps challenge the algorithms under a different set of conditions or parameters, similarly culminates in a set of data denoting the response times, also captured in milliseconds. The results from this separate analytical run are graphically mapped out in Fig. 8. This dual-format display of results ensures a thorough interpretation of the data, aiding in the comparative scrutiny essential for holistic understanding.

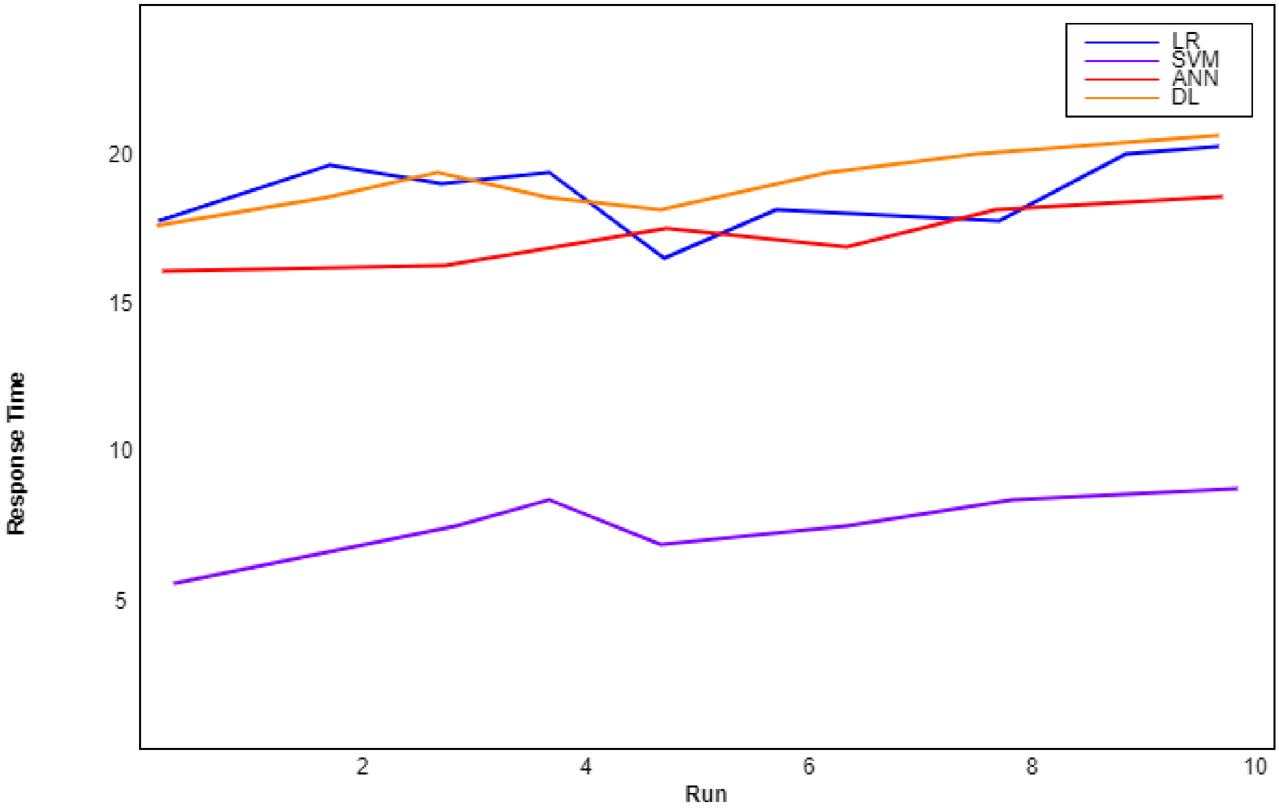


Fig. 7. Obtained results using machine learning methods in first case study.

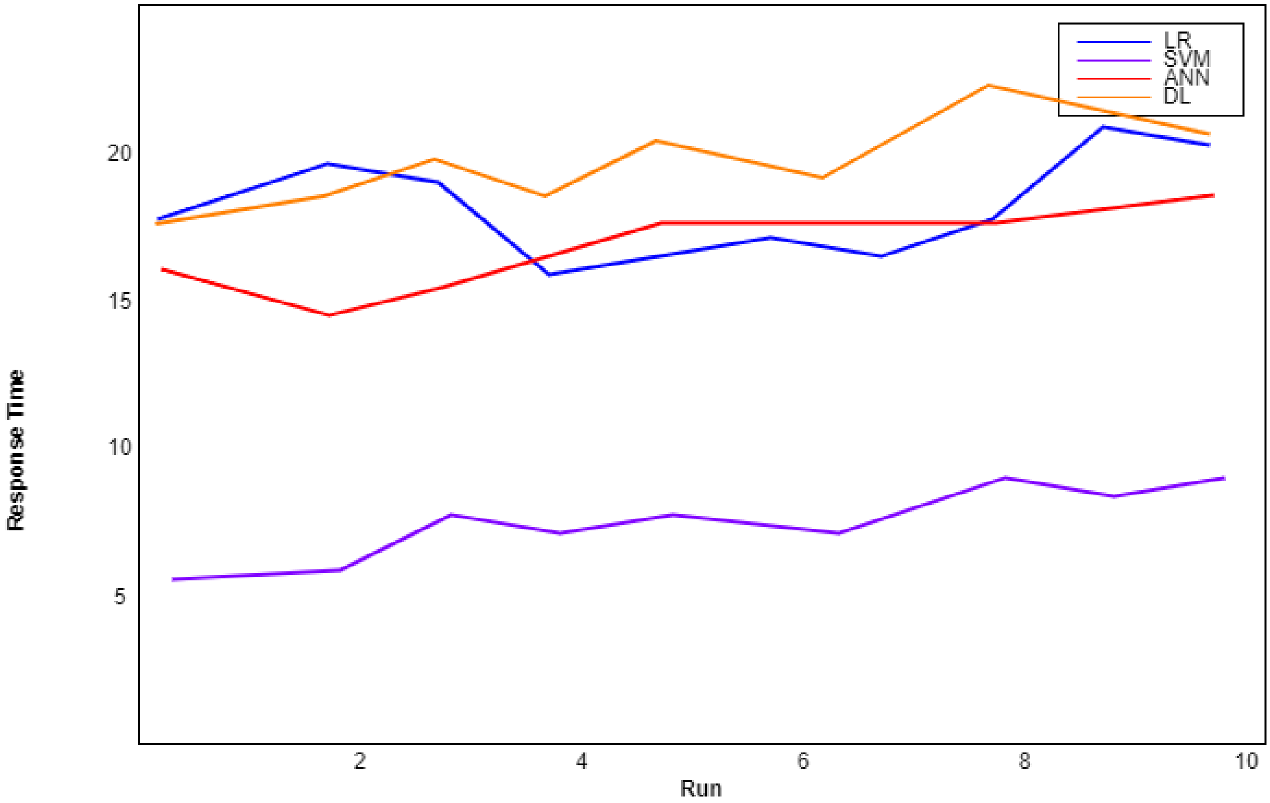


Fig. 8. Obtained results using machine learning methods in second case study.

Furthermore, an average response time was computed for each scenario, providing a consolidated metric that accounts for variance and anomalies by diluting them across multiple runs. This average was derived by executing each algorithm 100 consecutive times and then calculating the mean response time from the total accumulated data. This process, therefore, ensures the reliability and consistency of the results, acknowledging that a single run can be susceptible to anomalous external influences [57].

Such a meticulous approach to data representation not only underscores the rigorousness of the testing environment but also provides readers with a clear, unambiguous representation of each algorithm's performance. This systematic presentation and analysis are crucial for informed decision-making, further research, and practical applications of these algorithms in real-world scenarios. Fig. 9 demonstrates ROC curves of the applied machine learning methods for software defined networks.

Performance assessment stands as a pivotal component in the realm of machine learning (ML), particularly when it involves the critical evaluation of classification problems. One of the paramount tools in this evaluative arsenal is the Receiver Operating Characteristic (ROC) curve, an instrumental diagnostic graph that elucidates the competency of a classification model by displaying the trade-off between sensitivity (true positive rate) and specificity (false positive rate) across various thresholds [58].

The essence of the ROC curve lies in its capacity to illustrate the probability curve for distinct classes, thereby providing insightful commentary on the model's predictive acumen. It offers a graphical representation that is integral in comprehending how adeptly a model discerns between classes, underlining the probabilities that it correctly identifies true positives and true negatives. This finesse in estimation is paramount, as it significantly influences the decision-making processes.

In the context of the research at hand, the ROC curves were employed as an evaluative measure for distinct scenarios, each formulated to test the mettle of specific algorithms: Support Vector Machine (SVM) with regularization parameters C = 1 and C = 100, Artificial Neural Networks (ANN), and Deep Learning (DL). These scenarios, differentiated possibly by their unique datasets, conditions, or objectives, necessitated an assessment methodology that would impartially and accurately reflect the performance of each algorithm.

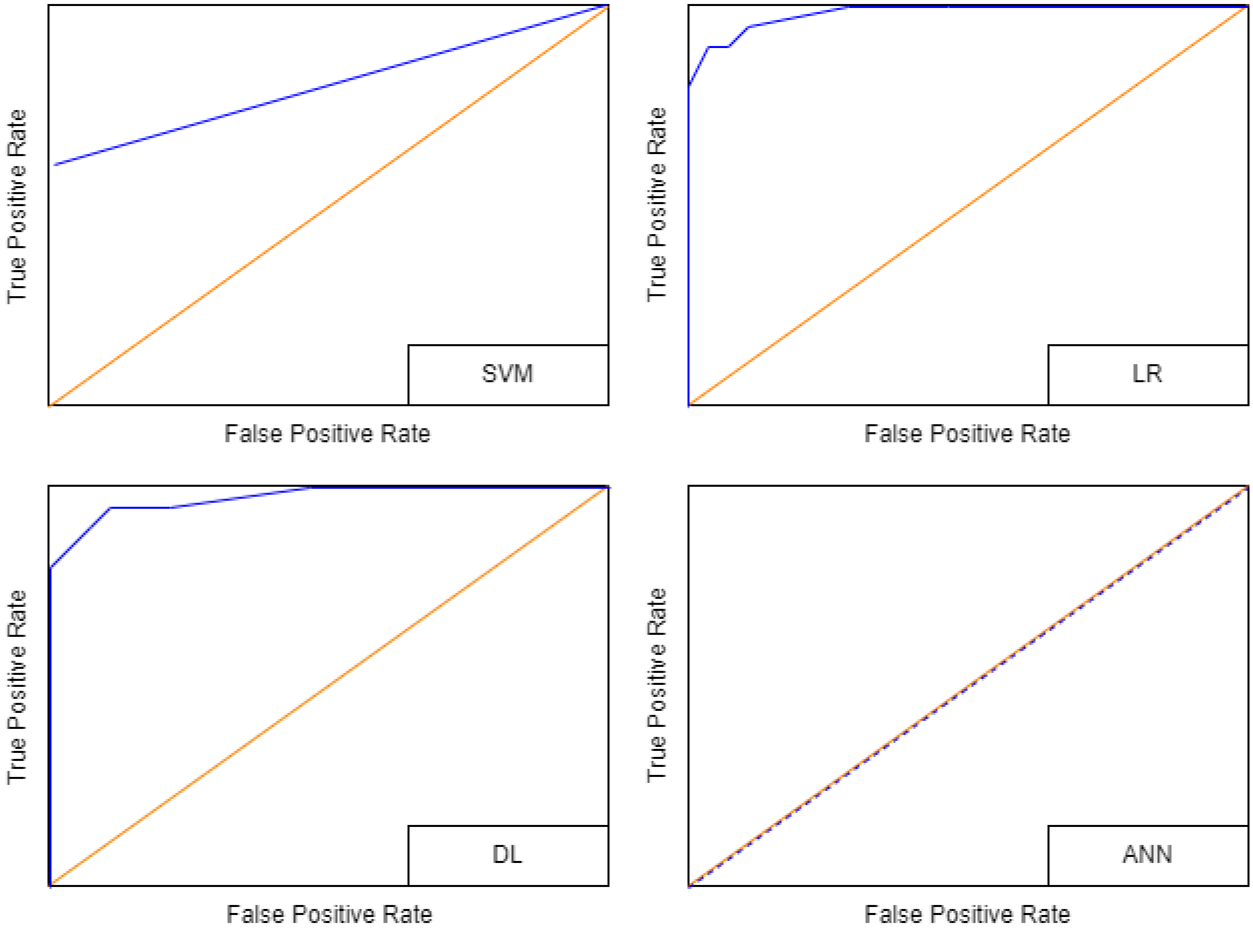


Fig. 9. ROC curves of different methods in applying machine learning for SDN.

The results of this meticulous evaluation are visually encapsulated in Fig. 9. Each figure corresponds to a different scenario and presents the ROC curves for the algorithms, thereby showcasing a comparative analysis of their performance in terms of sensitivity and specificity. This side-by-side portrayal of the ROC curves is instrumental in not only understanding the individual performance nuances of SVM, ANN, and DL within each scenario but also in drawing insightful inferences from their comparative capabilities [59].

Such analytical representations are invaluable, as they transcend mere numerical accuracy and delve into the model's ability to maintain robustness across varying class distributions and thresholds. This depth of analysis is imperative for the practical application of machine learning models, ensuring they are vetted not just on the scaffold of accuracy, but on their holistic performance and reliability in diverse operational environments.

V. Discussion

In the complex and rapidly evolving field of machine learning applied within Software-Defined Networks (SDNs), the current research initiates a nuanced dialogue, offering insights into a pioneering approach that leverages advanced algorithms for enhanced load balancing [60]. This discussion section delves into the multifaceted aspects of the study, critically analyzing the advantages, confronting the challenges, acknowledging limitations, and envisioning future trajectories.

A. Advantages

The integration of machine learning in SDN heralds a transformative phase in network management, primarily by introducing predictive analytics, automation, and adaptability. Firstly, the model's ability to predict traffic patterns and potential security threats stands as a testament to its predictive prowess, which traditional networking models lack [61]. By analyzing and learning from historical network data, the system can anticipate future loads, enabling proactive adjustments.

Moreover, automation in load balancing, facilitated by the employed machine learning techniques, significantly reduces the need for human intervention, thereby minimizing human errors and operational costs. It also frees up valuable human resources for more complex, creative tasks, enhancing productivity [62].

Furthermore, the adaptability inherent in machine learning models ensures that the system evolves with the changing network scenarios and traffic conditions. This dynamism stands in stark contrast to the static nature of traditional network configurations and is integral to managing the unpredictable, ever-changing demands on modern networks.

B. Challenges

However, the path to such integration is strewn with challenges. One of the foremost is the complexity involved in training the models. The chosen machine learning algorithms require substantial computational power and time, especially as the system scales, and the data volume grows.

Data security and privacy are other critical concerns. The necessity to collect, store, and analyze vast amounts of network data opens potential vulnerabilities for confidential information, necessitating robust security protocols that could increase overhead costs [63].

Moreover, the system's efficiency is contingent on the quality of the data fed into it. Inconsistent, incomplete, or biased data can skew the machine learning models' outcomes, leading to inaccurate predictions and suboptimal load balancing.

C. Limitations

Reflecting on the limitations of the current study underscores areas for improvement. The research predominantly focused on theoretical modeling and controlled environments, which may not fully simulate the unpredictability and heterogeneity of real-world network traffic [64].

The study's efficacy in a live environment, subjected to various unforeseen variables and security threats, remains unverified. Thus, the research, though sound in its theoretical foundation, necessitates empirical testing within a practical, operational network setting.

Another notable limitation is the selection of features and parameters for the machine learning models [65]. While the research justifies the chosen variables, there remains a question of whether other overlooked parameters could influence the load balancing process. Also, the algorithms' performance has been evaluated in isolation, without considering the potential synergies or conflicts that could arise from an integrated, multi-algorithm approach.

D. Future Directions

Addressing these limitations and challenges outlines the future trajectory of this research area. One immediate step is the transition from a controlled environment to real-world testing. Implementing the proposed model within operational SDNs will provide invaluable insights into its practical applicability and resilience, especially in the face of security threats and anomalous traffic patterns [66].

Considering the exponential growth in global data traffic, future studies must also explore models that can seamlessly scale, without a corresponding exponential demand on resources [67]. Such research could delve into more resource-efficient machine learning algorithms or hybrid models that combine the strengths of multiple algorithms to achieve more effective load balancing with lesser computational demand.

The aspect of security, given its centrality in network operations, also calls for dedicated exploration. Future research endeavors could focus on developing integrated security protocols within the machine learning models, ensuring that data privacy and network security are intrinsic to the system rather than afterthoughts.

Additionally, given the foundational role of data in machine learning, subsequent studies should investigate comprehensive, unbiased, and representative data collection methods. Enhancing the quality of input data can significantly bolster the accuracy and reliability of the predictions and decisions made by the machine learning models.

On a broader spectrum, there lies immense potential in interdisciplinary research, particularly integrating behavioral sciences with machine learning. Understanding the human aspects of network usage can add another layer of sophistication to predictive models, potentially leading to more intuitive, user-centric network management.

In conclusion, while the current study marks a significant stride towards revolutionizing SDN through machine learning, it also sets the stage for further, more nuanced exploration and innovation. The journey ahead, though demanding, holds the promise of networks that are not just smarter and more efficient but are also more in tune with the human elements they serve. Through collaborative, interdisciplinary, and forward-thinking research, the vision of truly intelligent, secure, and user-focused networks is an achievable horizon.

VI. Conclusion

The journey through this research, from conceptual frameworks to analytical discussions, reflects a profound exploration of integrating machine learning into Software-Defined Networking (SDN) to enhance load balancing. As we draw conclusions, it's imperative to encapsulate the essence of our findings and their implications for future scientific inquiry and practical application in the networking sphere.

This study marked a significant advancement by demonstrating that machine learning algorithms could revolutionize the way network resources are managed, optimizing the distribution of data loads across various pathways. By employing sophisticated algorithms, we unveiled the potential to predict network congestions, dynamically adjust to traffic changes, and improve overall efficiency and user experience. This paradigm shift from traditional methods accentuates a move towards more autonomous, self-sufficient systems capable of sophisticated decision-making processes, essential in the burgeoning era of digital transformation and the Internet of Things (IoT).

However, the research also highlighted critical challenges and limitations, from the complexities of algorithm training and data security concerns to the practical applicability of the proposed model outside simulated environments. These challenges are not terminuses but instead signposts indicating areas requiring further exploration, refinement, and innovation.

Looking forward, the implications of this research are both broad and profound. They suggest an imminent need for robust, real-world testing and the potential for interdisciplinary approaches that could further enrich these technological advances. The prospects of enhanced security measures, scalability considerations, and user-centric adaptations also present exciting, necessary trajectories.

In conclusion, this study does not represent an end but a beginning. It serves as a catalyst for continued exploration and dialogue in the realms of machine learning, networking, and beyond. The confluence of these fields holds significant promise for creating more resilient, efficient, and intelligent networks, poised to support the ever-evolving demands of future digital landscapes.

References

[1] Hamdan, M., Hassan, E., Abdelaziz, A., Elhigazi, A., Mohammed, B., Khan, S., ... & Marsono, M. N. (2021). A comprehensive survey of load balancing techniques in software-defined network. Journal of Network and Computer Applications, 174, 102856.

[2] Omarov, B., & Altayeva, A. (2018, January). Towards intelligent IoT smart city platform based on OneM2M guideline: smart grid case study. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 701-704). IEEE.

[3] Muhammad, T. (2022). A Comprehensive Study on Software-Defined Load Balancers: Architectural Flexibility & Application Service Delivery in On-Premises Ecosystems. INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, 6(1), 1-24.

[4] Rahman, A., Islam, J., Kundu, D., Karim, R., Rahman, Z., Band, S. S., ... & Kumar, N. (2023). Impacts of blockchain in software-defined Internet of Things ecosystem with Network Function Virtualization for smart applications: Present perspectives and future directions. International Journal of Communication Systems, e5429.

[5] Jurado-Lasso, F. F., Marchegiani, L., Jurado, J. F., Abu-Mahfouz, A. M., & Fafoutis, X. (2022). A survey on machine learning software-defined wireless sensor networks (ml-SDWSNS): Current status and major challenges. IEEE Access, 10, 23560-23592.

[6] Wu, J., Dong, M., Ota, K., Li, J., & Yang, W. (2020). Application-aware consensus management for software-defined intelligent blockchain in IoT. IEEE Network, 34(1), 69-75.

[7] Yazdinejad, A., Parizi, R. M., Dehghantanha, A., & Choo, K. K. R. (2020). P4-to-blockchain: A secure blockchain-enabled packet parser for software defined networking. Computers & Security, 88, 101629.

[8] A. Altayeva, B. Omarov, H.C. Jeong and Y.I. Cho, "Multi-step face recognition for improving face detection and recognition rate", Far East Journal of Electronics and Communications, vol. 16, no. 3, pp. 471-491, 2016

[9] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019). Applying face recognition in video surveillance security systems. In Software Technology: Methods and Tools: 51st International Conference, TOOLS 2019, Innopolis, Russia, October 15–17, 2019, Proceedings 51 (pp. 271-280). Springer International Publishing.

[10] Rawal, B. S., Manogaran, G., Singh, R., Poongodi, M., & Hamdi, M. (2021, June). Network augmentation by dynamically splitting the switching function in SDN. In 2021 IEEE International Conference on Communications Workshops (ICC Workshops) (pp. 1-6). IEEE.

[11] Latif, S. A., Wen, F. B. X., Iwendi, C., Li-Li, F. W., Mohsin, S. M., Han, Z., & Band, S. S. (2022). AI-empowered, blockchain and SDN integrated security architecture for IoT network of cyber physical systems. Computer Communications, 181, 274-283.

[12] Wang, Y., Shang, F., Lei, J., Zhu, X., Qin, H., & Wen, J. (2023). Dual-attention assisted deep reinforcement learning algorithm for energy-efficient resource allocation in Industrial Internet of Things. Future Generation Computer Systems, 142, 150-164.

[13] Cao, B., Sun, Z., Zhang, J., & Gu, Y. (2021). Resource allocation in 5G IoV architecture based on SDN and fog-cloud computing. IEEE Transactions on Intelligent Transportation Systems, 22(6), 3832-3840.

[14] Keshari, S. K., Kansal, V., Kumar, S., & Bansal, P. (2023). An intelligent energy efficient optimized approach to control the traffic flow in Software-Defined IoT networks. Sustainable Energy Technologies and Assessments, 55, 102952.

[15] Poornima, E., Muthu, B., Agrawal, R., Kumar, S. P., Dhingra, M., Asaad, R. R., & Jumani, A. K. (2023). Fog robotics-based intelligence transportation system using line-of-sight intelligent transportation. Multimedia Tools and Applications, 1-29.

[16] Razdan, S., & Sharma, S. (2022). Internet of medical things (IoMT): Overview, emerging technologies, and case studies. IETE technical review, 39(4), 775-788.

[17] Kazmi, S. H. A., Qamar, F., Hassan, R., Nisar, K., & Chowdhry, B. S. (2023). Survey on joint paradigm of 5G and SDN emerging mobile technologies: Architecture, security, challenges and research directions. Wireless Personal Communications, 1-48.

[18] Amiri, Z., Heidari, A., Navimipour, N. J., & Unal, M. (2023). Resilient and dependability management in distributed environments: A systematic and comprehensive literature review. Cluster Computing, 26(2), 1565-1600.

[19] Banafaa, M., Shayea, I., Din, J., Azmi, M. H., Alashbi, A., Daradkeh, Y. I., & Alhammadi, A. (2023). 6G mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities. Alexandria Engineering Journal, 64, 245-274.

[20] Ray, P. P., & Kumar, N. (2021). SDN/NFV architectures for edge-cloud oriented IoT: A systematic review. Computer Communications, 169, 129-153.

[21] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. Computers, Materials & Continua, 74(3).

[22] Mughaid, A., AlZu'bi, S., Alnajjar, A., AbuElsoud, E., Salhi, S. E., Igried, B., & Abualigah, L. (2023). Improved dropping attacks detecting system in 5g networks using machine learning and deep learning approaches. Multimedia Tools and Applications, 82(9), 13973-13995.

[23] Rahman, A., Islam, M. J., Montieri, A., Nasir, M. K., Reza, M. M., Band, S. S., ... & Mosavi, A. (2021). Smartblock-sdn: An optimized blockchain-sdn framework for resource management in iot. IEEE Access, 9, 28361-28376.

[24] Ribeiro, D. A., Melgarejo, D. C., Saadi, M., Rosa, R. L., & Rodríguez, D. Z. (2023). A novel deep deterministic policy gradient model applied to intelligent transportation system security problems in 5G and 6G network scenarios. Physical Communication, 56, 101938.

[25] Javanmardi, S., Shojafar, M., Mohammadi, R., Persico, V., & Pescapè, A. (2023). S-FoS: A secure workflow scheduling approach for performance optimization in SDN-based IoT-Fog networks. Journal of Information Security and Applications, 72, 103404.

[26] Kashef, M., Visvizi, A., & Troisi, O. (2021). Smart city as a smart service system: Human-computer interaction and smart city surveillance systems. Computers in Human Behavior, 124, 106923.

[27] Qu, Y., Wang, Y., Ming, X., & Chu, X. (2023). Multi-stakeholder's sustainable requirement analysis for smart manufacturing systems based on the stakeholder value network approach. Computers & Industrial Engineering, 177, 109043.

[28] Bourechak, A., Zedadra, O., Kouahla, M. N., Guerrieri, A., Seridi, H., & Fortino, G. (2023). At the Confluence of Artificial Intelligence and Edge Computing in IoT-Based Applications: A Review and New Perspectives. Sensors, 23(3), 1639.

[29] Imam-Fulani, Y. O., Faruk, N., Sowande, O. A., Abdulkarim, A., Alozie, E., Usman, A. D., ... & Taura, L. S. (2023). 5G Frequency Standardization, Technologies, Channel Models, and Network Deployment: Advances, Challenges, and Future Directions. Sustainability, 15(6), 5173.

[30] Abou El Houda, Z., Hafid, A. S., & Khoukhi, L. (2023). Mitfed: A privacy preserving collaborative network attack mitigation framework based on federated learning using sdn and blockchain. IEEE Transactions on Network Science and Engineering.

[31] Sheng, M., Zhou, D., Bai, W., Liu, J., Li, H., Shi, Y., & Li, J. (2023). Coverage enhancement for 6G satellite-terrestrial integrated networks: performance metrics, constellation configuration and resource allocation. Science China Information Sciences, 66(3), 130303.

[32] Sutradhar, S., Karforma, S., Bose, R., & Roy, S. (2023). A Dynamic Step-wise Tiny Encryption Algorithm with Fruit Fly Optimization for Quality of Service improvement in healthcare. Healthcare Analytics, 3, 100177.

[33] Al-Turjman, F., Zahmatkesh, H., & Shahroze, R. (2022). An overview of security and privacy in smart cities' IoT communications. Transactions on Emerging Telecommunications Technologies, 33(3), e3677.

[34] Mahi, M. J. N., Chaki, S., Ahmed, S., Biswas, M., Kaiser, M. S., Islam, M. S., ... & Whaiduzzaman, M. (2022). A review on VANET research: Perspective of recent emerging technologies. IEEE Access, 10, 65760-65783.

[35] Omarov, B., Altayeva, A., & Cho, Y. I. (2017). Smart building climate control considering indoor and outdoor parameters. In Computer Information Systems and Industrial Management: 16th IFIP TC8 International Conference, CISIM 2017, Bialystok, Poland, June 16-18, 2017, Proceedings 16 (pp. 412-422). Springer International Publishing.

[36] Zhou, H., Zheng, Y., Jia, X., & Shu, J. (2023). Collaborative prediction and detection of DDoS attacks in edge computing: A deep learning-based approach with distributed SDN. Computer Networks, 225, 109642.

[37] Zhang, J., Liu, Y., Li, Z., & Lu, Y. (2023). Forecast-assisted service function chain dynamic deployment for SDN/NFV-enabled cloud management systems. IEEE Systems Journal.

[38] Priyadarshini, R., & Barik, R. K. (2022). A deep learning based intelligent framework to mitigate DDoS attack in fog environment. Journal of King Saud University-Computer and Information Sciences, 34(3), 825-831.

[39] Das, S. K., Benkhelifa, F., Sun, Y., Abumarshoud, H., Abbasi, Q. H., Imran, M. A., & Mohjazi, L. (2023). Comprehensive review on ML-based RIS-enhanced IoT systems: basics, research progress and future challenges. Computer Networks, 224, 109581.

[40] Mubarakali, A., Durai, A. D., Alshehri, M., AlFarraj, O., Ramakrishnan, J., & Mavaluru, D. (2023). Fog-based delay-sensitive data transmission algorithm for data forwarding and storage in cloud environment for multimedia applications. Big Data, 11(2), 128-136.

[41] Liu, D., Li, Z., & Jia, D. (2023). Secure distributed data integrity auditing with high efficiency in 5G-enabled software-defined edge computing. Cyber Security and Applications, 1, 100004.

[42] Omarov, B., Suliman, A., & Tsoy, A. (2016). Parallel backpropagation neural network training for face recognition. Far East Journal of Electronics and Communications, 16(4), 801-808.

[43] Gong, J., & Rezaeipanah, A. (2023). A fuzzy delay-bandwidth guaranteed routing algorithm for video conferencing services over SDN networks. Multimedia Tools and Applications, 1-30.

[44] Altayeva, A. B., Omarov, B. S., Aitmagambetov, A. Z., Kendzhaeva, B. B., & Burkitbayeva, M. A. (2014). Modeling and exploring base station characteristics of LTE mobile networks. Life Science Journal, 11(6), 227-233.

[45] Shahraki, A., Abbasi, M., Piran, M. J., & Taherkordi, A. (2021). A comprehensive survey on 6G networks: Applications, core services, enabling technologies, and future challenges. arXiv preprint arXiv:2101.12475.

[46] Ahmed, A. A., Malebary, S. J., Ali, W., & Barukab, O. M. (2023). Smart traffic shaping based on distributed reinforcement learning for multimedia streaming over 5G-VANET communication technology. Mathematics, 11(3), 700.

[47] Sahoo, S. K., Mudligiriyappa, N., Algethami, A. A., Manoharan, P., Hamdi, M., & Raahemifar, K. (2022). Intelligent trust-based utility and reusability model: Enhanced security using unmanned aerial vehicles on sensor nodes. Applied Sciences, 12(3), 1317.

[48] Mahmoodi Khaniabadi, S., Javadpour, A., Gheisari, M., Zhang, W., Liu, Y., & Sangaiah, A. K. (2023). An intelligent sustainable efficient transmission internet protocol to switch between User Datagram Protocol and Transmission Control Protocol in IoT computing. Expert Systems, 40(5), e13129.

[49] Fatemidokht, H., Rafsanjani, M. K., Gupta, B. B., & Hsu, C. H. (2021). Efficient and secure routing protocol based on artificial intelligence algorithms with UAV-assisted for vehicular ad hoc networks in intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems, 22(7), 4757-4769.

[50] Saba, T., Rehman, A., Sadad, T., Kolivand, H., & Bahaj, S. A. (2022). Anomaly-based intrusion detection system for IoT networks through deep learning model. Computers and Electrical Engineering, 99, 107810.

[51] Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., ... & Amira, A. (2023). AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Artificial Intelligence Review, 56(6), 4929-5021.

[52] Sangaiah, A. K., Javadpour, A., Ja'fari, F., Pinto, P., Zhang, W., & Balasubramanian, S. (2023). A hybrid heuristics artificial intelligence feature selection for intrusion detection classifiers in cloud of things. Cluster Computing, 26(1), 599-612.

[53] Javed, A. R., Shahzad, F., ur Rehman, S., Zikria, Y. B., Razzak, I., Jalil, Z., & Xu, G. (2022). Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects. Cities, 129, 103794.

[54] Alhayani, B., Kwekha-Rashid, A. S., Mahajan, H. B., Ilhan, H., Uke, N., Alkhayyat, A., & Mohammed, H. J. (2023). 5G standards for the Industry 4.0 enabled communication systems using artificial intelligence: perspective of smart healthcare system. Applied nanoscience, 13(3), 1807-1817.

[55] Tursynova, A., & Omarov, B. (2021, November). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. In 2021 16th International Conference on Electronics Computer and Computation (ICECCO) (pp. 1-4). IEEE.

[56] Zhang, Q., Li, C., Huang, Y., & Luo, Y. (2023). Effective multi-controller management and adaptive service deployment strategy in multi-access edge computing environment. Ad Hoc Networks, 138, 103020.

[57] bin Salleh, R., Koubaa, A., Khan, Z., Khan, M. K., & Ali, I. (2023). Data plane failure and its recovery techniques in SDN: A systematic literature review. Journal of King Saud University-Computer and Information Sciences.

[58] Xu, X., Li, H., Xu, W., Liu, Z., Yao, L., & Dai, F. (2021). Artificial intelligence for edge service optimization in internet of vehicles: A survey. Tsinghua Science and Technology, 27(2), 270-287.

[59] Khan, A. A., Laghari, A. A., Rashid, M., Li, H., Javed, A. R., & Gadekallu, T. R. (2023). Artificial intelligence and blockchain technology for secure smart grid and power distribution Automation: A State-of-the-Art Review. Sustainable Energy Technologies and Assessments, 57, 103282.

[60] Strinati, E. C., Alexandropoulos, G. C., Sciancalepore, V., Di Renzo, M., Wymeersch, H., Phan-Huy, D. T., ... & Denis, B. (2021, June). Wireless environment as a service enabled by reconfigurable intelligent surfaces: The RISE-6G perspective. In 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit) (pp. 562-567). IEEE.

[61] Haque, A. B., Bhushan, B., & Dhiman, G. (2022). Conceptualizing smart city applications: Requirements, architecture, security issues, and emerging trends. Expert Systems, 39(5), e12753.

[62] Ghiasi, M., Niknam, T., Wang, Z., Mehrandezh, M., Dehghani, M., & Ghadimi, N. (2023). A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future. Electric Power Systems Research, 215, 108975.

[63] Ali, O., Abdelbaki, W., Shrestha, A., Elbasi, E., Alryalat, M. A. A., & Dwivedi, Y. K. (2023). A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. Journal of Innovation & Knowledge, 8(1), 100333.

[64] Mori, S., Mizutani, K., & Harada, H. (2023). Software-Defined Radio-Based 5G Physical Layer Experimental Platform for Highly Mobile Environments. IEEE Open Journal of Vehicular Technology, 4, 230-240.

[65] Neubauer, M., Reiff, C., Walker, M., Oechsle, S., Lechler, A., & Verl, A. (2023). Cloud-based evaluation platform for software-defined manufacturing: Cloud-basierte Evaluierungsplattform für Software-defined Manufacturing. at-Automatisierungstechnik, 71(5), 351-363.

[66] Kube, A. R., Das, S., & Fowler, P. J. (2023). Fair and efficient allocation of scarce resources based on predicted outcomes: implications for homeless service delivery. Journal of Artificial Intelligence Research, 76, 1219-1245.

[67] Hashem, I. A. T., Usmani, R. S. A., Almutairi, M. S., Ibrahim, A. O., Zakari, A., Alotaibi, F., ... & Chiroma, H. (2023). Urban Computing for Sustainable Smart Cities: Recent Advances, Taxonomy, and Open Research Challenges. Sustainability, 15(5), 3916.

A Comprehensive Review of Healthcare Prediction using Data Science with Deep Learning

Asha Latha Thandu, Pradeepini Gera
Department of Computer Science and Engineering, Koneru Laksmaiah Education Foundation,
Vaddeswaram, Andhra Pradesh, 500302, India

***Abstract*—Data science in healthcare prediction technology can identify diseases and spot even the smallest changes in the patient's health factors and prevent the diseases. Several factors make data science crucial to healthcare today the most important among them is the competitive demand for valuable information in the healthcare systems. The data science technology along with Deep Learning (DL) techniques creates medical records, disease diagnosis, and especially, real-time monitoring of patients. Each DL algorithm performs differently using different datasets. The impacts on different predictive results may be affects overall results. The variability of prognostic results is large in the clinical decision-making process. Consequently, it is necessary to understand the several DL algorithms required for handling big amount of data in healthcare sector. Therefore, this review paper highlights the basic DL algorithms used for prediction, classification and explains how they are used in the healthcare sector. The goal of this review is to provide a clear overview of data science technologies in healthcare solutions. The analysis determines that each DL algorithm have several negativities. The optimal method is necessary for critical healthcare prediction data. This review also offers several examples of data science and DL to diagnose upcoming trends on the healthcare system.**

Keywords—Data science; deep belief network; healthcare; sparse auto encoder; deep learning

I. Introduction

Many studies have been conducted over the years on how to enhance the management and administration activities of the health sector and especially, healthcare offered to its patient [1]. Currently, the need of healthcare data is growing at an exponential rate in the healthcare system. From this point of view, the deployment of technology that is capable of being used in a creative way for the organization to help it achieve its goals is critical [2-5]. More preventive treatment options are becoming possible due to the use of health data analytics, especially predictive analytics. Despite access to a large amount of data, the healthcare industry lacks actionable knowledge that can be used to make predictions [6-10]. Despite its abundance, this is due to the fact that that health data is basically complex and fragmented [11]. Critical care, which is part of the health sector, faces the problem of increasing population and economic pressures, due to which it is difficult for most people to get the appropriate treatments. When talking to Intensive Care Unit (ICU) their condition often changes with every movement [12-14]. Likewise, with the advancement of technology in the healthcare sector an expectations of the patients are increasing but due to rising inflation, the required services cannot be provided. The main problem is to provide better, more effective care [15-17].

A data science solution for the analysis of healthcare data can help save patient lives and improve our quality of life [18-20]. Data science deals with several topics, such as data management and analysis, make correct decisions to improve the operation or system services (for instance, healthcare and transportation systems) [21-23]. In addition, with some very innovative and insightful techniques for displaying big data post-analysis, it is now easier to understand how any complex system works [24]. Due to the complexity of the healthcare system and operations, healthcare data are frequently fragmented. For example, different hospitals are only allowed to view clinical data of patients belonging to their specific patient groups [25]. These documents include highly Private Health Information (PHI) about specific individuals. This section examines the state-of-the-art in healthcare prediction using six deep representative architectures. Table I provides comparison of existing surveys with our survey.

The analysis shows that none of the studies carried out data science technology in healthcare prediction. Thus, this survey focuses on data science technologies in healthcare prediction. The contribution of this review is explained as follows:

- This review describes importance of data science technology in healthcare prediction in detail.
- Several deep learning technologies and applications used data science technology are analyzed for healthcare prediction.
- This survey helps researchers to diagnose the potential challenges in healthcare industry.

A. Data Collection

The searching keywords which are used in the above-mentioned databases like: “Healthcare Prediction” is demanded in every search abstract. It is a usual technique and consuming time as well. Also, searched for various synonyms and related keywords which meets the review outcomes such as, “Deep learning technology using data science technology”, “Applications in data science”, “Disease prediction using data science technology”, “healthcare monitoring” and “predictive analysis”. Again, refined the query to meet the particular outputs and again applied on Abstract and Title of the research article.

TABLE I. COMPARISON OF EXISTING SURVEYS

Author	Year	Description	Advantages	Drawbacks
Wang Y et al. [77]	2018	Analyzed benefits of big data based on information technology infrastructure, managerial, organizational, operational, and strategic areas	This study employs a clear vision about how healthcare organizations are using big data analytics	The limitation of this study is source of data because of the IT adoption lags behind other industries in healthcare sector
Shirazi S et al. [73]	2019	Discussed data mining approaches and algorithms in healthcare domain	Classifying data mining papers regarding unsupervised and supervised learning algorithms	Didn't give a clear vision of the applications of data mining technologies.
Bohr A and Memarzadeh K [78]	2020	Discussed major applications in artificial intelligence	Detailed research on applications directly related to healthcare and applications across the healthcare value chain including ambient assisted living and drug development	There is no detail about the different types of deep learning techniques
Li W et al. [79]	2021	Examined machine learning applications for big data in the healthcare sector	This survey examines basic big data concepts and relationship between big data and IoT	This review didn't provide detail about disease prediction using ML techniques like covid-19, mental health
Ours	-	To offer an overview of data science technologies in healthcare prediction	Detailed review about various deep learning technologies and applications using data science technology	-

On searched articles, implements a quality assessment procedure following inclusion and exclusion constraints. Derived 3050 articles depends on the common keywords from several journals from different sources, such as Springer, ACM, IEEE Xplore, and Science Direct. Further, excluded some research papers that are not related to this study based on the heading. Fig. 1 shows flow chart of article selection procedure.

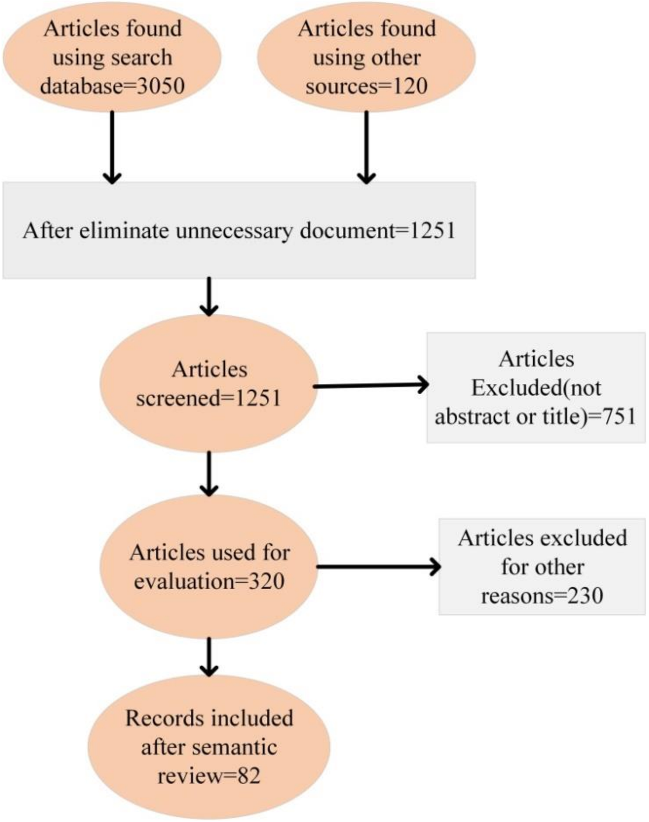

Articles found using search database=3050
Articles found using other sources=120
After eliminate unnecessary document=1251
Articles screened=1251
Articles Excluded(not abstract or title)=751
Articles used for evaluation=320
Articles excluded for other reasons=230
Records included after semantic review=82

Fig. 1. Flowchart of article selection process.

II. HEALTHCARE PREDICTION REVIEW METHODOLOGY USING VARIOUS DEEP LEARNING CLASSIFIERS

In this section, healthcare prediction review methodology using deep learning classifiers are explained. Fig. 2 represents the taxonomy of the Healthcare prediction review.

A. Healthcare Prediction using Data Science

1) Data acquisition: HealthData.gov, Big Cities Health Inventory Data Platform, Chronic Disease Data, Human Mortality Database, Mental Disorders Datasets, MHealth Dataset, Medicare Provider Utilization and Payment Data, Life Science Database Archive, and WHO (World Health Organization) are some of the general datasets [26] used in this study to gather information on healthcare. The "precision medicine initiative" relates to healthcare. It sought to map the human genomes of one million United states residents, identify specific genetic defects that underlie a specific disease in a population, and effectively direct the development of drugs that are capable of precisely addressing a subcategory of molecular problems distributed by patients with a particular illness. Clinical, genetic features, and pathological can be included in IBM Watson for Healthcare, which can then provide standardized therapeutic paths and personalized therapy suggestions based on those features. By using DL algorithms, the enclitic firm (San Francisco, CA, USA) improves diagnostic performance in minimum time and in decreased costs using healthcare images (like MRIs and X-rays). Another excellent application is google flu trends, which uses monitoring data from laboratories around the US to forecast influenza-like disease than twice of doctor's visits. Major research institutes, centers, and funding organizations have been able to invest in this field due to the significant role of these technologies in clinical and medical research. Along with health information systems, data science and DL techniques can be used to enhance healthcare management systems to meet the following objectives: cost reduction, fewer hospitalizations and shorter lengths of stays, prevention of fraudulent activity, classification in disease patterns, strong health insurance, and more efficient use of healthcare resources. The upcoming sections provide multiple application instances based on the various types of biomedical data, including biomedical time signals, biomedical images, and other biomedical information from wearable system, lab findings, and genomics. Modern biomedical equipment generates electrical signals from skin-mounted sensors, the features of which depend on the position of the sensor. These signals are an invaluable source of information for identifying and diagnosing diseases. By utilizing physiological signals like ElectroMyoGrams (EMG), ElectroEncephaloGrams (EEG), ElectroOculoGrams (EOG), and ElectroCardioGrams (ECG), DL can generate reliable applications.

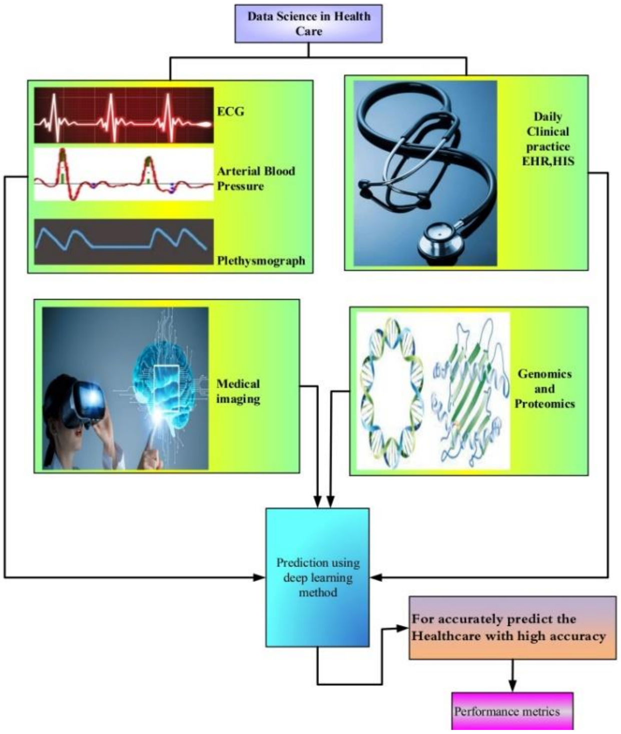

Data Science in Health Care
ECG
Arterial Blood Pressure
Plethysmograph
Daily Clinical practice EHR,HIS
Medical imaging
Genomics and Proteomics
Prediction using deep learning method
For accurately predict the Healthcare with high accuracy
Performance metrics

Fig. 2. Taxonomy of the healthcare prediction review.

Recently, the importance of data from EHRs has increased. These records contain substantial volumes of formless text including physical exams, medical lab results, operation notes, and discharge that are still difficult to approach. To support clinical decision-making system, natural language processing depending on deep learning is used to gather important data from the text. For instance, an application is used to identify healthy and patients with peripheral arterial disease by extracting necessary text data related to the condition from narrative clinical notes. Measuring the semantic similarity of medical concepts is a challenge that is critical to several strategies in information retrieval and medical informatics. A Big Data (BD) framework has more recently been used to analyze the issue of preventing and treating Acute Coronary Syndrome (ACS). A multitask learning architecture based on adversarial learning techniques was presented to detect significant adverse cardiac events from ACS patients' EHR data. This adversarial learning frame work outperforms single-subtype focused models in terms of average prediction, when the three subcategories of ACS are included in the detection. Regardless of the subtypes, the parameter distributions of common features are comparable across true-positive and false-positive samples as well as between true-negative and false-negative samples. This occurs due to the small amount of patient samples are available in the patients record. This demonstrates the potential benefits of Big Data (BD) approaches, resides not only in the need to extract a significant amount of data from heterogeneous HER, but also in the development of DL procedure that can: (1) Deal with the diversity (in images, text, radiological reports, etc.); (2) Reduce over-fitting and enhance the generalization capability; (3) Deal with the uncertainty that the missing information presents (common in EHR). These difficulties are not the latest ones, and there are still unresolved ones in the ML community. Therefore, this work is reviewed using a several deep learning techniques and applications.

B. Deep Learning Methods for Health Care Prediction

In order to capture hierarchical relationship incorporated in deep features, DL has emerged as one of the primary study issues in the field of prognostics due to the fast growth of computing infrastructure. The deep network structure with numerous layers stacked in the network to completely capture the relevant data from the initial input data. In numerous domains, including image identification, audio recognition, and natural language processing, DL models have attracted significant interest and achieved notable successes. However, in the area of healthcare prediction, it has not yet been completely utilized. Six sample deep architectures—Deep Belief Network (DBN) [27], Convolutional Neural Network (CNN) [28], Recurrent Neural Network (RNN) [29], Long Short-Term Memory (LSTM) [30], Auto-encoder [31], and Sparse auto encoder [32]—were primarily the focus on the published research on DL. Based on these six exemplary deep architectures, this section aims to examine existing techniques.

The deep learning methods are, DBNs, CNN, RNN, LSTM, Auto encoders (AEs), and sparse Autoencoders (SAEs). Fig. 3 shows deep learning classifiers used in Health care prediction.

1) Deep belief network: A stack of Restricted Boltzmann Machines (RBMs) called the DBN consists of higher-order BMs as well as feature-identifying units on a single layer in BMs. The greedy layer-wise learning procedure of RBMs may pre-train the approach in an unsupervised manner with no limitations on the volume of training data.

A DBN is a generative statistical approach that can learns deep representations of input data. It stimulates the combined distribution of the hidden layers and the observable data. The RBM, an energy-based generative approach consisting of input layers, hidden layers, and symmetric networks between them, is one example of a basic, unsupervised network known as a DBN. In this situation, the incoming RBM uses the current hidden layer as its input layer.

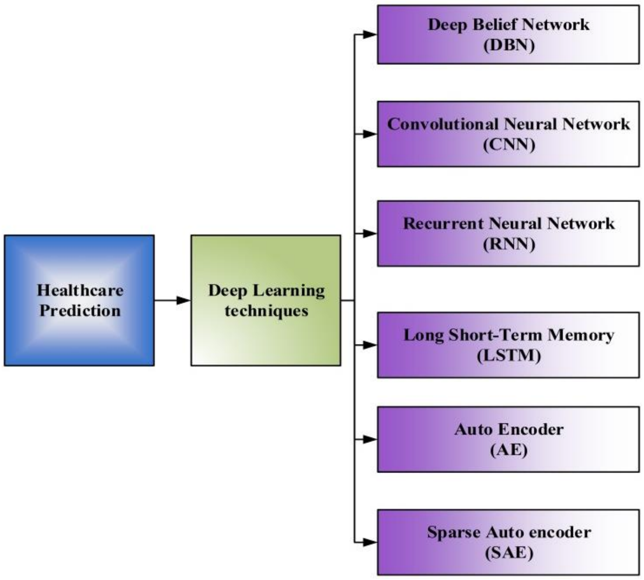

Healthcare Prediction
Deep Learning techniques
Deep Belief Network (DBN)
Convolutional Neural Network (CNN)
Recurrent Neural Network (RNN)
Long Short-Term Memory (LSTM)
Auto Encoder (AE)
Sparse Auto encoder (SAE)

Fig. 3. Deep learning techniques used in Healthcare prediction.

The backpropagation technique may also be used to adjust the whole network. A rapid, layer-by-layer unsupervised architecture is created as a result of this composition, and it has since emerged as the most powerful DL algorithms. The capacity of DBN to recreate the input in an unsupervised manner is one of its standout features. For this reason, it has been used to carry out effective unsupervised feature learning in the fields of healthcare and transportation. For instance, the study used DBN to identify health care in input data. DBN is thought to learn poorly for anomalous samples while rebuilding the input, which often leads to significant reconstruction errors. Setting an error threshold enables the efficient detection of aberrant data.

Fig. 4 depicts CNN's framework structure. Every CNN has input, output, and hidden layers. There are several function layers in the invisible layer. Most functional layers are composed of convolutional, pooling, fully connected, and normalized layers. The foundation of CNN is the convolutional layer. The input amount is calculated by each filter in the convolution process using a dot operation. The attributes are not explicitly extracted from the data by the convolutional layer. The Eq. (1) is utilized to compute it.

$$(f*h)(t)\underline{\underline{def}}\int_{-\infty}^{\infty} f(\Gamma)h(t-\Gamma)d\Gamma \tag{1}$$

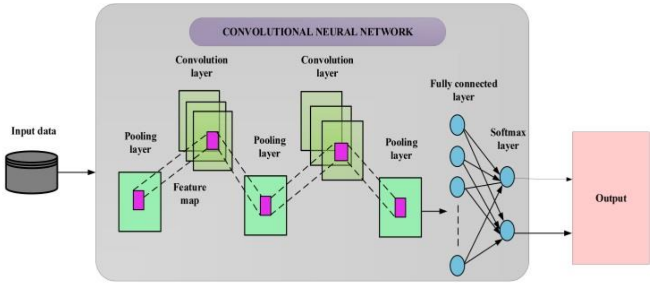


Fig. 4. Framework of CNN.

where, $(f*h)(t)$ are convolution of two functions f and h at time domain t. $\underline{\underline{def}}$ represents the definition sign, $h(t-\Gamma)$ denotes the weighting function and $f(\Gamma)$ represents the input function. Convolution is the multiplication of the function that has been moved and inverted in the equation above.

The pooling layer for CNN includes both local and global pooling. The assessment of healthcare data is lowered when output neurons and a particular neuron are merged. Through pooling, the maximum and average value of neurons are determined.

This fully connected layer is also called a thick layer. It is situated in CNN's last section. The neurons in this connected layer are entirely connected to all of the activation layers from the preceding stage. Each neuron is composed of a single layer that is joined to another layer by a substantial layer.

This activation layer enables the network to plot the nonlinear function versus the complicated action. This is divided into two categories: saturated activation and non-saturated activation. The activation is referred to as being saturated when there are no output constraints; otherwise, it is referred to as non-saturated. The CNN framework successfully learns the structural properties of the data by taking into consideration the spatial and temporal information that is included in the data. As a consequence, calculation time is decreased, and classification accuracy is improved.

2) Recurrent neural network: An RNN is a complex design that can interpret dynamic input because it has feedback links from hidden or output layers to the previous layer. The common design for incoming data, including temperature-time series data and pressure-time, is sequential data. Although RNN is the popular sequential modelling approaches, it has limits on long-term Remaining Useful Life (RUL) forecasts due to the inability of the trained network's weights to identify trends as a result of weight updates made when each input pattern was presented. Because of this, researchers have created the LSTM, which overcomes problems with long-term time dependence by employing input gates, forget gates, and output gates to manage information flow. In many sequential applications, the RNN and its variation, the LSTM networks, have gained considerable popularity. In recent years, healthcare prediction researchers have begun to investigate the potential of RNN, particularly LSTM. The extended kalman filter, truncated back propagation via time gradient computation, and evolutionary algorithms make up the RNN training algorithm. RNN are employed in this method to classify the health prediction data. An explanation of the classifier is given below.

There are various input layers (Y1, Y2), numerous hidden levels, and only one output layer in the recurrent neural network. The construction of RNN is seen in Fig. 5.

Healthcare detection is categorized using an RNN classifier. Healthcare input data are computed by RNN. There are several hidden layers to it. For the purpose of receiving output healthcare data, the network is connected to the proper layer. The input parameter for the network is weights and biases. The irrelevant data in each layer may be computed using the soft sign activation function as shown in Eq. (2),

$$f(y)=\frac{y}{1+|y|} \tag{2}$$

In the above equation, input features of the network are denoted as y.

The hidden layer's output is given in Eq. (3),

$$l_t = g(Z_l y_s + V_l l_{s-1} + a_l) \tag{3}$$

In the above equation, Y_s represents input data, l_t represents hidden layer, Z_l and V_l represents network weighted value, a_l represents biases, l_{s-1} denotes the previous state of the hidden layer and $g(\ )$ represents activation function.

By employing the Gaussian activation function, the determined outcome is given to the input of next layer to detect the healthcare data. Gaussian output is calculated as in Eq. (4),

$$Gaussion = e^{-y^2} \tag{4}$$

where, $Gaussion$ denotes the output of gaussian function, y be the variable. Then the total output is computed as in Eq. (5).

$$x_s = f(Z_x l_s + a_x) \tag{5}$$

In above equation, x_s represents the network output, l_s represents hidden layer, Z_x represents network weighted value, a_x represents biases and $f(\ )$ represents activation function. It results higher precision with lower execution time.

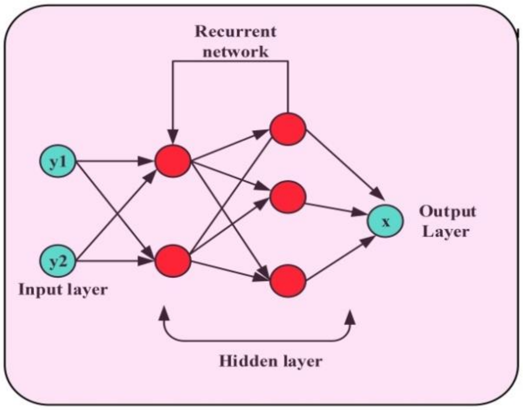


Fig. 5. Construction of RNN.

3) Long short-term memory: The LSTM network is a category of DL network created especially for linear data processing. An advantage of LSTM is their ability to retain both short-term and long-term values. Each cell in an LSTM unit has inputs, outputs, and forget gates. These three gates control the flow of information. By looking at the health prediction data, this LSTM categorizes the healthcare data in data science technology.

Numerous additional researchers have focused their attention on LSTM applications. LSTM, a kind of RNN for modelling long-term dependencies, was developed to solve the difficulties of time-series data. Like ordinary RNNs, LSTMs contain a memory for replicating the hidden layer activation patterns. Data processing involves hidden layer activations replicated iteratively.

The field of healthcare is paying more and more attention to LSTM models. According to sensor information such as blood pressure, temperature, and the results of lab tests, an LSTM model was employed as an example to diagnose healthcare data in a hospital setting. Similar to this, an LSTM model was applied to forecast examination outcomes based on prior measures. DeepCare is an LSTM-based system that is used to forecast future healthcare outcomes and infer the current sickness condition. A growing corpus of research has also focused on employing LSTMs to extract detailed data from medical texts like scientific publications, such as names of medications or medical occurrences.

4) Auto-encoder: The input data's encoder and decoder phases are rebuilt so that the Auto Encoder (AE) can learn a different representation of the information. As a result, it is frequently utilized for network pretraining. The stacked SAE, which combines multilayer AE such as denoising AE and SAE, is widely used Deep Neural Network (DNN) methods for handling the data. The main use of AE models in deep health monitoring is defect diagnosis. There are very few direct uses of AE in healthcare prediction in the literature; instead, AE is often employed to derive degradation features.

The unsupervised mode is applied when a hidden layer recreates the input layer in auto-encoders. Dimensions are allocated, in contrast to RNN, which incorporates weights and bias from the input layer to the hidden layer. The non-linear transformation function is utilized to calculate the stimulation of the hidden layer, which has lower dimensions compared to the input layer. Meanwhile, the input displays a dominating structure when the hidden layer size is decreased. To understand the identity function, non-linearity function should not be included, and the hidden layer and input layer dimensions should remain unchanged.

There are two subcategories of autoencoders. The first subcategory is denoising auto-encoder, and the approach is suggested to stop learning from becoming a simple problem. In this case, the input is rebuilt using noise that has been corrupted. Stacked auto-encoders are another form; they are created by stacking one auto-encoder layer on top of another. Each layer is trained separately to anticipate the output in healthcare applications, and the entire network is tweaked using supervised learning techniques. The architecture of AEs is symmetrical, with the equal number of nodes in both the input layer and the output layer. The employment of AE has advantages such as dimensionality reduction and feature learning. However, there are certain problems with dimension reduction and feature extraction in AE. The code of AE's emphasis on minimizing data relationship loss results in the loss of several crucial data relationships. Its classification equation is given in Eq. (6),

$$\beta, \chi = \arg\min_{\beta,\chi} E\left(Y, \left(\chi \circ \beta\right) Y\right) \tag{6}$$

where, β is represented as the encoder, χ is represented as the decoder, E is represented as the error between input and output, Y is represented as the input data, $\arg\min_{\beta,\chi}$ Minimum augmented of encoder and decoder.

5) Sparse auto encoder: The SAE is simply applying sparse constraints to the AE code. The unseen layer has more neurons than the input layer and output layer combined. While some of these neurons have a propensity to zero, others are similar to the input neurons. These are a few advantages of sparsity: The Sparse Auto Encoder's Structure is depicted in Fig. 6.

- The model's anti-noise performance is enhanced over that of the original AE because it facilitates the extraction of key input characteristics.
- Sparseness is easier to understand and explain because it is satisfied by the majority of real-world circumstances.

The classification of healthcare prediction using SAE equation is given inEq. (7),

$$SAE = E + \alpha \sum_{i} KL^{'}\left(\rho \| \hat{\rho}_i\right) \tag{7}$$

where, SAE represents the output of sparse autoencoder, $KL^{'}$ is represented as the kullback-leibler divergence, $KL^{'}$ divergence is known as the benchmark function that measures the effectiveness of different two disseminations are ρ and $\hat{\rho}$, ρ are represented as the anticipated neuronal network activity level, $\hat{\rho}$ indicates the average level of activation with i^{th} neuron, α maintains the weight parameter.

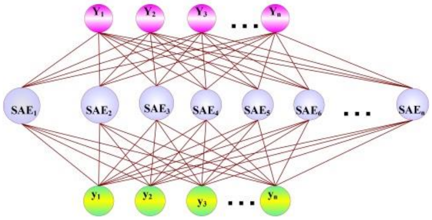


Fig. 6. Structure of Sparse auto encoder.

C. Comparison of Various Mechanisms on Healthcare Prediction

In this section, comparison of different DL techniques is reviewed based on the strength and weakness of several DL models used for healthcare prediction. This review also considers accuracy term for comparison to determine the effectiveness of each model. Table II provides comparison of previously published research works advantages and their disadvantages.

TABLE II. REVIEW OF EXISTING WORKS ADVANTAGE AND DISADVANTAGE

Reference	Aim	Used technique	DL technique	Accuracy	Positives	Negatives	Scope for improvements
Lu et al. [33]	Cardiovascular disease prediction	DBN	DBN	91.26%	Good stability	Minimum accuracy rate	Applying implemented model in terms of depth learning for cardiovascular prediction
Ali et al. [34]	Heart disease prediction	Optimally Configured and Improved DBN (OCI-DBN	DBN	94.61%	Supports doctors to take efficient decisions	Not consider time complexity	Time complexity for the suggested technique will be computed because it most necessary factor in healthcare
Elkholy et al. [35]	Chronic kidney disease prediction	Modified DBN	DBN	98.5%	Beneficial for clinical decision making	Not support for unbalanced dataset	-
Javeed at al. [36]	Human behavior recognition model	Sustainable Healthcare Pattern Recognition (SPHR)	DBN	93.33% and 92.50%	Autonomous feature extraction reduces the dependency on domain expert	It gives lower accuracy for static activities	To perform the suggested model on complex activities
Pan et al. [37]	Heart disease prediction	Enhanced Deep learning assisted Convolutional Neural Network (EDCNN)	CNN	94.9%	Supports specialists to forecast information about heart patient using cloud platforms wherever in the world	Not applied in a real-world	Performance is further improved using feature selection techniques
Chung et al. [38]	Healthcare recommendations	CNN	CNN	90.1%	A dynamic cluster mechanism was suggested as well as prediction accuracy improved based	This recommendations system not suitable for symbolic	Further work can be extended to support symbolic knowledge expansion framework

					on user environment, which changes over time and has dynamic components rather than static components in terms of environmental factors	knowledge expansion framework	using AI techniques
Gaur et al. [39]	Covid-19 detection	Deep CNN	CNN	92.93%	Appropriate for mobile applications	Less amount of X-ray images was used for implementation	Several deep learning approaches and models will be implemented for further research work
Younis et al. [40]	Brain tumor prediction	Visual Geometry Group (VGG16)	CNN	98.14%	The work uses MRI images to classify brain tumors and to support in making fast, effective and correct decisions	Different types of cancer attacks cannot be identified.	Future work will be considered to differentiate brain tumor affected region from unaffected area precisely
Kim et al. [41]	Prediction of five chronic diseases	Character-RNN (Char-RNN)	RNN	77.6%, 82.6%, 80.6%, 82.5%, 96.5%	Efficiency for several chronic disease prediction	This mechanism was applied only to the Korean peoples	Need to apply the suggested model to different ethnicities and lifestyle habits
Zhu et al. [42]	Prediction of future glucose levels	Dilated Recurrent Neural Network (DRNN)	RNN	-	This approach performs better than existing forecasting algorithms	Inaccurate prediction.	In future, the suggested method will be embedded with IoT app
Feng et al. [43]	Healthcare prediction for football players	Smart football player health prediction approach	RNN	81%	Better reliability	Less prediction rate	-
Ma et al. [44]	Parkinson’s chronic disease prediction	Self-attention and RNN	RNN	93.55%	Better Parkinson’s prediction skills	Not implemented in real-time	-
Carrillo-Moreno et al. [45]	Glucose predictor	LSTM	LSTM	-	To offer best prediction to patient, the suggested classifier validates several prediction times and input dimensions	This system considers only three parameters for glucose prediction	Adding few more parameters in the prediction system and computing the importance of these parameters for glucose prediction
Said et al. [46]	Covid-19 detection	Bi-directional LSTM (Bi-LSTM)	LSTM	-	Better detection performance	Experiment was conducted on data from the Qatar	Establishing lockdown relaxation scenario and analyze the impact of the relaxation
Mohamed et al. [47]	RNA mutation prediction	Seq2seq LSTM	LSTM	98.9% and 96.9%	The obtained outcomes illustrate the possibility of applying the LSTM network to RNA and DNA sequences in solving other sequencing problems	Improved computational complexity	-
Algarni et al. [48]	Human emotion recognition	Stacked Bi-LSTM	LSTM	99.45%, 96.87% and 96.68%	The model’s performance results help to make precise medical decisions	The Bi-LSTM provides complexity in weight initialization process	Applying new algorithms on several datasets to validate efficiency in emotion recognition
Mallick et al. [49]	Cancer detection using brain MRI images	Deep Wavelet Autoencoder (DWA)	AE	96%	Achieved good results	Reliability problems	Integrating other variation of AE with DNN
Mahendran et al. [50]	Signal compression	priority-based convolutional AE	AE	-	Zero construction error	Less performance	Improving performance by focusing on the CNN framework
Mansour et al. [51]	Covid-19 prediction	Unsupervised DL based Variational AE	AE	98.7% and 99.2%	Good classification performance	Not suitable for e-healthcare applications	The suggested model embedded with IoT and cloud-based environment
Khamparia et al. [52]	Chronic kidney disease classification	Stacked AE	AE	100%	Good accuracy	Not tested on large datasets	Testing larger datasets for disease classification.

Ebiaredoh-Mienye et al. [53]	Disease prediction	Enhanced SAE	SAE	98%, 97% and 91%	Efficient feature learning and better performance	It doesn't consider several parameters such as computational speed, classification time and etc.	The implemented model incorporated with decision support system to help doctors
Mienye et al. [54]	Heart disease prediction	Stacked SAE	SAE	97.3% and 96.1%	Significant importance in classification performance	Efficiency issues	Focusing on other stacking variation of AE to observe effects on classification performance
Aslam et al. [55]	Breath analysis	Stacked SAE	SAE	98.7% and 97.3%	More reliable	Not accurate	-
Gayathri et al. [56]	Covid-19 prediction	Feed forward neural network	SAE	95.78%	Achieved good accuracy	Not suitable for multi-classification	Covid-19 prediction using other modalities like ultrasound and CT

D. Multimedia Healthcare Applications using Deep Learning Technologies

An integration of several media or several types of data from multiple devices like texts, images, videos or audios called multimedia or multimodal data. To enhance the performance of the application, complementary information can be extracted from each modality by extracting multimodal data. Modality means encode information in a particular way. Various perspectives of a physiological objects using multimodal data provide additional information that can complementary to the analysis. Table III provides multimedia healthcare applications using DL algorithms.

TABLE III. HEALTHCARE APPLICATIONS USING DL ALGORITHMS

Author	**Applications**	**DL technique**	**Multimedia data**	**Database**
Gaur L et al. [39]	Covid-19 classification	Deep CNN	Chest X-ray	Covid-19 radiography database and actualmed-covid-cheat x-ray dataset
Alhussein M and Muhammad G [80]	Voice pathology prediction	Parallel CNN	Voice signals	Saarbrucken Voice Database
Algarni M et al. [48]	Emotion recognition	Bi-LSTM	EEG signal	DEAP dataset
Mukherjee D et al. [81]	Human activity recognition	EnsemConvNet	Time series of data	WISDM dataset, MobiAct dataset, UniMiB SHAR dataset
Yu Z et al. [82]	Disease Prediction	Deep Factorization Machine	Patients medical history	2020 artificial intelligence challenge preliminary competition

E. Applications in Data Science Technology

Different applications of data science in healthcare prediction is explained in this section. This review considers different applications including speech therapy, disease detection, drug discovery, health monitoring, genomics and decision support system.

1) Speech therapy: Speech therapy supports children and adults affected with communication disorder to improve speech and languages. It supports with sound and voice production, early language skills, fluency and clarity. A speech therapist can use several types of therapy to support individuals with communication disorder related to fluency, speech, language, cognition, voice and swallowing. Mahmoud SS et al. (2020) [57] suggested assessment mechanism. Quadrature-based high-resolution time-frequency images with a CNN are used to detect the relationship between speech intensity and three speech intelligibility features in aphasic patients. The outcomes show a linear relationship with statistically significant correlations between the CNN model's normalized Truth-Class Output Functions (TCOA) and patients' pronunciation, tone scores, and fluency. Also, Bastanfard A et al. (2009) [58] proposed a new method that adopts image-based technique to combine visemes in persian by using coarticulation effect. The central frame was selected among various images for each phoneme defining various positions in different symbols. As a result of reconstruction, the weight value was established as criterion to compare viseme similarity. Experimental outcomes demonstrate the excellent precision and robustness of the suggested model.

2) Psychological prediction: This section explains specific DL algorithm used for disease detection including mental health, neurodevelopment disorder, and covid-19.

a) Mental health: In the real-word, the one of the most important and complex concern is auto diagnosis of mental health conditions. The mental health affects the way people behave, think and feel when they cooperate with world around them. Additionally, mental health issues are becoming a leading disability, contributing greatly to the global burden of disease. Du C et al. (2021) [59] designed a DL based Mental Health Monitoring System (DL-MHMS) for academy students. By using EEG signals, this suggested method used the effective CNN to categorize status of the mental health as normal, negative and positive. Zeberga K et al. (2023) [60] used Bidirectional Encoder Representations in Transformers (BERT) for excellently and efficiently recognizing anxiety and depression based posts by monitoring the context and semantic meaning of the words. Additionally, the knowledge distillation approach was proposed to transfer knowledge from a large pretrained model to a smaller model, which enhances accuracy. In last stage, BERT with Bi-LSTM and word2vec efficiently detects depression and anxiety symptoms.

b) Neurodevelopment disorders: Neurodevelopmental Disorders (NDs) affects brain functions as well as neurological developments, which causes problems in cognitive, social and emotional functioning. Some of the NDs are dyslexia, Autism Spectrum Disorder (ASD), and Attention Deficit Hyperactivity Disorder (ADHD). Sewani H et al. (2020) [61] offered an efficient prognosis of ASD using deep neural network particularly for children. This model integrates unsupervised learning, an AE and supervised DL using CNNs. The suggested approach performs better based on several validation and assessment measures. This model was tested on only 1112 rs-fMRI images. Minoofam SA et al. (2022) [62] suggested an adaptive reinforcement learning framework called RALF automatically generates content for students with dyslexia by Cellular Learning Automata (CLA). First, RALF creates samples of online alphabet as a simple form. The CLA system learns every instructions of character formation asynchronously using a reinforcement learning cycle. Then, generates persian words algorithmically. This stage determines the position of the character, cursiveness of the letters and the cell's response to the environment. At last, RALF uses embedded word-generation approach to generate long-texts and sentences. Spaces between words are obtained using CLA neighboring states. The developed model offers many tests and games to enhance word pronunciation and people's learning performance.

3) Covid-19 detection: More than five lakhs people died in India due to the covid-19. Generally, covid-19 often causes respiratory symptoms such as a cold, pneumonia or flu. Covid-19 can attack individuals' lungs and respiratory system. Panwar H et al. (2020) [63] offered a DL algorithm for rapidly prediction of covid-19 cases depending on CT scan and X-ray images because the X-ray images offers necessary information in the covid-19 recognition. The model can identify covid-19 positive cases in less than two seconds, which is quicker than RT-PCR test. Kogilavani SV et al. (2022) [64] proposed several CNN designs like VGG-19, Densenet121, MobileNet, NASNet, Xception and EfficientNet to identify covid-19. This model didn't detect covid-19 affected areas in the lungs.

4) Drug discovery: The importance of the drug discovery procedure is the advancement of novel drugs with potential interactions along with therapeutic targets. Drug discovery is depending on the conventional method, which focuses on holistic treatment. The medical communities of the world began to use the allopathic method to treatment and recovery in the last century. This change has led to victory in fighting against diseases, but has resulted in high healthcare burden and drug costs. Xiong Z et al. (2019) [65] introduced a novel graph neural network framework named Attentive FP to represent molecular, that used attention model to learn from suitable drug discovery datasets. The suggested approach achieves good performance on diverse datasets and that its learning is interpretable. The suggested Attentive FP automatically learns intramolecular interactions from specific tasks, helps derive chemical insights directly from data beyond human perception. ul Qamar MT et al. (2020) [66] analyzed the viral three-Chymotrypsin-Like cysteine protease enzymes, which is necessary for coronavirus lifecycle and controls its replication. This mechanism has been constructs 3D homology model by analyzing Chymotrypsin-Like cysteine protease sequences and screened it against a medicinal plant library comprising 32,297 potential anti-viral phytochemicals/ traditional Chinese medicinal compounds. The analysis demonstrates that the first nine hits the process of drug development to combat covid.

5) Healthcare monitoring: Data science serves important role in IoT. These devices consist of wearable devices that monitor heartbeat, temperature and medical parameters of the patients. Then, collected data is analyzed using data science. Based on the analytical results, doctors can able to monitor patient's circadian cycle, their BP, and calorie intake. Ali F et al. (2021) [67] recommended BD analytics engine depending on data mining approaches, ontologies and Bi-LSTM. The data mining approaches are employed to preprocess the healthcare data and dimensionality reduction. The suggested ontologies are used to learn semantic knowledge about entities and features, and their relationships in Blood Pressure (BP) and diabetes domain. Finally, Bi-LSTM classifier is suggested to categorize effects from the drug side and abnormal states in individuals. The obtained results of the suggested method correctly handle big amount of data and enhances classification accuracy and prediction of drug side effect. M Abd El-Aziz R et al. (2022) [68] suggested IoT supported health monitoring system, which improves the data processing efficiency with the rapid adoption of cloud computing and expands data access in the cloud. The collected information from the real-time environment was stored in the cloud for data science processing. In this, improved pigeon optimization was suggested to combine the stored data in the cloud that supported for enhancing prediction rate. Following that, feature extraction and selection are performed with the help of optimal feature selection approaches. To categorize human healthcare, Backtracking Search-Based DNN (BS-DNN) was suggested.

6) Genomics: Genomics is the study of the sequencing and analysis of genes. A genome contains DNA and all the genes of an organism. Since the compilation of the Human Genome Project, research has progressed rapidly and embedded itself in the data science and big data fields. Nasir MU et al. (2022) [69] suggested CNN based AlexNet to predict genome multi-class disorder for developing Advance Genome Disorder Prediction Model (AGDPM). This model was capable of genome disorder prediction and it uses large amount of data to processes patient's genome disorder data. The suggested prediction system improves biomedical system based on predict genetic disorders and reduces high mortality rates.

7) Decision support system: The goal of a Decision Support System (DSS) is to enhance healthcare delivery through improving clinical decisions with targeted clinical knowledge, patient data and other health related information.

Malmir B et al. (2017) [70] presented a DSS called Fuzzy Expert System(FES) to support doctors for better decision making in medical diagnosis. This suggested model conducts a cross-sectional study to gather information about diseases by asking clinicians on all signs based on diseases. According to this information, then fuzzy rule-based system with the necessary symptoms necessary signs based on the suspected disease was developed. To prove effectiveness of the suggested method, two case studies conducted on kidney stone and kidney infection. Khiabani SJ et al. (2022) [71] developed DSS in terms of neural network and statistical process control charts for identifying and control of Myocardial Infarction (MI) and continuous observing of patient BP. A group of patients was used to prove the suggested system's efficiency. The outcomes validate that the suggested model can detect MI by the parameter of accuracy and precision.

III. Results and Discussions

A review of data science and deep learning techniques for healthcare prediction with high accuracy is discussed in this section.

Fig. 7 illustrates the various deep learning techniques contributes to covid-19 prediction by the parameter of accuracy. The analysis demonstrate that the AE performs over other deep learning techniques including SAE, and CNN. But it is important to note that each work uses different datasets to prove its effectiveness. Correspondingly, the work based on AE [51] predicts covid-19 using chest x-ray images. It is necessary to diagnose covid-19 using other modalities such as CT, MRI and so on.

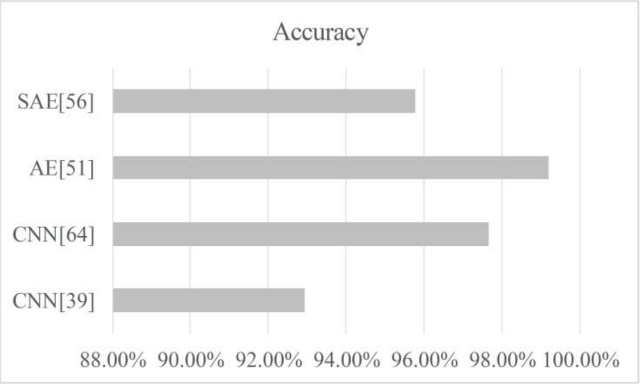


Fig. 7. Comparison of covid-19 detection using reviewed techniques.

Table IV shows the comparison of DL techniques in terms of precision and sensitivity for heart disease prediction. The analysis shows that the EDCCNN technique occurs better precision than other techniques. For the sensitivity analysis, the authors [54] uses optimized feature learning techniques, this supports to achieves a highest sensitivity of 100%.

Even though AI-based DL are efficiently predicting the data, some of the classifiers reduce the accuracy due to their limited data size, high dimensionality, efficient feature selection technique, model generalizability, and clinical implementation. The explanations for these limitations are given as follows:

1) *Limited data size:* One of the challenges facing this study was inadequate data for training the classifier. The less input data size allows a number of the training set, which reduces the accuracy of the presented methods. Hence, to improve the accuracy of the training samples and to train a large number of datasets new methods are used, which performs better than previous classifiers [72].

2) *High dimensionality:* High dimensionality is another problem faced by the previous methods. The input data set consists of number of data with a high number of features. The current methods face several high dimensionality issues for extracting the features. Hence, new feature-extracting methods are used to overcome these issues [73].

3) *Efficient feature selection technique:* Previously several feature selection and disease prediction methods were employed for predicting the disease in its early stage. But they are limited due to high computational complexity. To overcome this limitation, the most effective feature selection methods and pre-processing methods are used for predicting the data's higher accuracy [74].

4) *Model generalizability:* For improving the prediction results, a change in the research is needed based on the model's generalizability. Generalizability is used to analyze the results from highly populated situations with prediction methods. Previously there were several prediction methods for predicting the patient data in a single site. Nowadays, it is required to predict the patient data in multiple sites, and while improving the predicting data in multiple sites the model's generalizability is enhanced [75].

5) *Clinical implementation:* Finally, AI-based ML and DL methods have provided good results for predicting disease in DS in healthcare predictions. But still, this method faces issues in practical implementations with clinics that are not supported. In the current work, these AI methods are required to validate the clinical setting for assisting the doctor in confirming the findings and decisions [76].

TABLE IV. Comparison of Different Performance Metrics based on Heart Disease Prediction

Methods	Precision	Sensitivity
DBN [33]	-	-
OCI-DBN [34]	93.55%	96.03%
EDCNN [37]	99%	97.51%
Stacked SAE [54]	94.8%	100%

Many clinical elements are involved in EHR-based systems. There are millions of data points available for this. It would not be easy to manage and regulate the whole data of millions of people. There are several critical challenges yet to be overcome:

- There was a lot of unorganized or inaccurate data, making it tough to gain a more profound knowledge of it.

- It is difficult to strike the right balance between the preservation of patient-centric data and the excellence and convenience of this information.
- Keeping data private, storing it efficiently, and transferring it requires a lot of workforces to ensure that these requirements are met continuously.
- Lack of language proficiency when handling data.

IV. CONCLUSION

This manuscript proposes a review of data science and healthcare prediction in data science technology in order to forecast healthcare with high accuracy is successfully predicted. In this, the healthcare prediction is classified using deep learning classifiers based on health issues such as lab reports, medical imaging and EHR. In this, the deep learning classifiers are CNNs, RNN, LSTM, RBMs, DBNs, AEs and SAE. In healthcare services, the analysis shows that the existing approaches are not efficient to handle big data. Existing and future smart healthcare systems requires examinations about the design considerations such as maintainability, performance, accuracy, scalability, cost, security, responsiveness, fault tolerance, and reliability. It's hoped that this review paper will help many scholars to improve their knowledge for their future research work. Future research is planned to review machine and deep learning techniques with an approach that optimises healthcare data in different environment like IoT, cloud and so on.

REFERENCES

[1] Ismail A, Abdlerazek S, El-Henawy IM (2020) Big data analytics in heart diseases prediction. J Theor APPL Inf Technol 98(11):15-9.

[2] Lee C, Luo Z, Ngiam KY, Zhang M, Zheng K, Chen G, Ooi BC, Yip WL (2017) Big healthcare data analytics: Challenges and applications. In: Khan S, Zomaya A, Abbas A (eds) Handbook of Large-Scale Distributed Computing in Smart Healthcare. Scalable Computing and Communications, Springer, Cham. https://doi.org/10.1007/978-3-319-58280-1_2.

[3] Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2018) Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 19(6):1236-1246.

[4] Krishnamoorthi R, Joshi S, Almarzouki HZ, Shukla PK, Rizwan A, Kalpana C, Tiwari B (2022) A novel diabetes healthcare disease prediction framework using machine learning techniques. J Healthc Eng 2022. https://doi.org/10.1155/2022/1684017.

[5] Alotaibi S, Mehmood R, Katib I, Rana O, Albeshri A. Sehaa (2020) A big data analytics tool for healthcare symptoms and diseases detection using Twitter, Apache Spark, and Machine Learning. Appl Sci 10(4):1398.

[6] Wang Y, Kung L, Byrd TA (2018) Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technol Forecast Soc Change 126:3-13.

[7] Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J (2019) A guide to deep learning in healthcare. Nat Med 25(1):24-29.

[8] Saleem TJ, Chishti MA (2019) Deep learning for Internet of Things data analytics. Procedia Comput Sci 163:381-390.

[9] Abedjan Z, Boujemaa N, Campbell S, Casla P, Chatterjea S, Consoli S, Costa-Soria C, Czech P, Despenic M, Garattini C, Hamelinck D. Data science in healthcare: Benefits, challenges and opportunities. Data Sci Healthc 3-8.

[10] Sarwar MU, Hanif MK, Talib R, Mobeen A, Aslam M (2017) A survey of big data analytics in healthcare. Int J Adv Comput Sci Appl 8(6).

[11] Alharthi H (2018) Healthcare predictive analytics: An overview with a focus on Saudi Arabia. J Infect Public Health 11(6):749-756.

[12] Palanisamy V, Thirunavukarasu R (2019) Implications of big data analytics in developing healthcare frameworks–A review. J King Saud Univ - Comput Inf Sci 31(4):415-425.

[13] Gruson D, Helleputte T, Rousseau P, Gruson D (2019) Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. Clinic Biochem 69:1-7.

[14] Raeesi Vanani I, Amirhosseini M (2021) IoT-based diseases prediction and diagnosis system for healthcare. Internet of Things for Healthcare Technologies, Springer, Singapore, pp 21-48.

[15] Khan ZF, Alotaibi SR (2020) Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective. J Healthc Eng 2020:1-5.

[16] Ismail A, Shehab A, El-Henawy IM (2019) Healthcare analysis in smart big data analytics: reviews, challenges and recommendations. Security in Smart Cities: Models, Applications, and Challenges, Springer, Cham, pp 27-45.

[17] Syed L, Jabeen S, Manimala S, Elsayed HA (2019) Data science algorithms and techniques for smart healthcare using IoT and big data analytics. Smart Techniques for a Smarter Planet, Springer, Cham, pp 211-241.

[18] Wang L, Alexander CA (2020) Big data analytics in medical engineering and healthcare: methods, advances and challenges. J Med Eng Technol 44(6):267-283.

[19] Rizwan A, Zoha A, Zhang R, Ahmad W, Arshad K, Ali NA, Alomainy A, Imran MA, Abbasi QH (2018) A review on the role of nano-communication in future healthcare systems: A big data analytics perspective. IEEE Access 6:41903-41920.

[20] Nayyar A, Gadhavi L, Zaman N (2021) Machine learning in healthcare: review, opportunities and challenges. Machine Learning and the Internet of Medical Things in Healthcare 23-45.

[21] Subramaniyan S, Regan R, Perumal T, Venkatachalam K (2020) Semi-supervised machine learning algorithm for predicting diabetes using big data analytics. Business Intelligence for Enterprise Internet of Things, Springer, Cham, pp. 139-149.

[22] Nancy AA, Ravindran D, Raj Vincent PD, Srinivasan K, Gutierrez Reina D (2022) IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning. Electronics 11(15):2292.

[23] Saheb T, Izadi L (2019) Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends. Telemat Inform 41:70-85.

[24] Sahoo AK, Pradhan C, Das H (2020) Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making. Nature inspired computing for data science, Springer, Cham, 201-212.

[25] Mehta N, Pandit A (2018) Concurrence of big data analytics and healthcare: A systematic review. Int J Med Inform 114:57-65.

[26] Bote-Curiel L, Munoz-Romero S, Gerrero-Curieses A, Rojo-Álvarez JL (2019) Deep learning and big data in healthcare: a double review for critical beginners. Appl Sci 9(11):2331.

[27] Wan JJ, Chen BL, Kong YX, Ma XG, Yu YT (2019) An early intestinal cancer prediction algorithm based on deep belief network. Sci Rep 9(1):1-3.

[28] Sampathkumar A, Tesfayohani M, Shandilya SK, Goyal SB, Shaukat Jamal S, Shukla PK, Bedi P, Albeedan M (2022) Internet of Medical Things (IoMT) and Reflective Belief Design-Based Big Data Analytics with Convolution Neural Network-Metaheuristic Optimization Procedure (CNN-MOP). Comput Intell Neuroscience 2022.

[29] Feng Q, Liu Y, Wang L (2021) Wearable device-based smart football athlete health prediction algorithm based on recurrent neural networks. J Healthc Eng 2021.

[30] Koç E, Türkoğlu M (2022) Forecasting of medical equipment demand and outbreak spreading based on deep long short-term memory network: the COVID-19 pandemic in Turkey. Signal Image Video Process 16(3):613-621.

[31] Huang G, Wang H, Zhang L (2022) Sparse-Coding-Based Autoencoder and Its Application for Cancer Survivability Prediction. Math Probl Eng 2022.

[32] Hannah S, Deepa AJ, Chooralil VS, BrillySangeetha S, Yuvaraj N, Arshath Raja R, Suresh C, Vignesh R, Srihari K, Alene A (2022) Blockchain-based deep learning to process IoT data acquisition in cognitive data. BioMed Res Int 2022.

[33] Lu P, Guo S, Zhang H, Li Q, Wang Y, Wang Y, Qi L (2018) Research on improved depth belief network-based prediction of cardiovascular diseases. J Healthc Eng 2018.

[34] Ali SA, Raza B, Malik AK, Shahid AR, Faheem M, Alquhayz H, Kumar YJ (2020) An optimally configured and improved deep belief network (OCI-DBN) approach for heart disease prediction based on Ruzzo–Tompa and stacked genetic algorithm. IEEE Access 8:65947-65958.

[35] Elkholy SM, Rezk A, Saleh AA (2021) Early prediction of chronic kidney disease using deep belief network. IEEE Access 9:135542-135549.

[36] Javeed M, Gochoo M, Jalal A, Kim K (2021) HF-SPHR: Hybrid features for sustainable physical healthcare pattern recognition using deep belief networks. Sustainability 13(4):1699.

[37] Pan Y, Fu M, Cheng B, Tao X, Guo J (2020) Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform. IEEE Access 8:189503-189512.

[38] Chung K, Jung H (2020) Knowledge-based dynamic cluster model for healthcare management using a convolutional neural network. Inf Technol Manag 21:41-50.

[39] Gaur L, Bhatia U, Jhanjhi NZ, Muhammad G, Masud M (2021) Medical image-based detection of COVID-19 using deep convolution neural networks. Multimed Syst. https://doi.org/10.1007/s00530-021-00794-6.

[40] Younis A, Qiang L, Nyatega CO, Adamu MJ, Kawuwa HB (2022) Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches. Appl Sci 12(14):7282.

[41] Kim C, Son Y, Youm S (2019) Chronic disease prediction using character-recurrent neural network in the presence of missing information. Appl Sci 9(10):2170.

[42] Zhu T, Li K, Chen J, Herrero P, Georgiou P (2020) Dilated recurrent neural networks for glucose forecasting in type 1 diabetes. J Healthc Inform Res 4:308-324.

[43] Feng Q, Liu Y, Wang L (2021) Wearable device-based smart football athlete health prediction algorithm based on recurrent neural networks. J Healthc Eng 2021:1-7.

[44] Ma B, Zhang F, Ma B (2021) Self-Attention-Guided Recurrent Neural Network and Motion Perception for Intelligent Prediction of Chronic Diseases. J Healthc Eng 2021.

[45] Carrillo-Moreno J, Pérez-Gandía C, Sendra-Arranz R, García-Sáez G, Hernando ME, Gutiérrez Á (2021) Long short-term memory neural network for glucose prediction. Neural Comput Appl 33:4191-4203.

[46] Said AB, Erradi A, Aly HA, Mohamed A (2021) Predicting COVID-19 cases using bidirectional LSTM on multivariate time series. Environ Sci Pollut Res. 28(40):56043-56052.

[47] Mohamed T, Sayed S, Salah A, Houssein EH (2021) Long short-term memory neural networks for RNA viruses mutations prediction. Math Probl Eng 2021:1-9.

[48] Algarni M, Saeed F, Al-Hadhrami T, Ghabban F, Al-Sarem M (2022) Deep learning-based approach for emotion recognition using electroencephalography (EEG) signals using Bi-directional long short-term memory (Bi-LSTM). Sensors 22(8):2976.

[49] Mallick PK, Ryu SH, Satapathy SK, Mishra S, Nguyen GN, Tiwari P (2019) Brain MRI image classification for cancer detection using deep wavelet autoencoder-based deep neural network. IEEE Access 7:46278-46287.

[50] Mahendran RK, Velusamy P, Pandian P (2021) An efficient priority-based convolutional auto-encoder approach for electrocardiogram signal compression in Internet of Things based healthcare system. Trans Emerg Telecommun Technol 32(1):e4115.

[51] Mansour RF, Escorcia-Gutierrez J, Gamarra M, Gupta D, Castillo O, Kumar S (2021) Unsupervised deep learning based variational autoencoder model for COVID-19 diagnosis and classification. Pattern Recognit Lett. 151:267-274.

[52] Khamparia A, Saini G, Pandey B, Tiwari S, Gupta D, Khanna A (2020) KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network. Multimed Tools Appl 79:35425-35440.

[53] Ebiaredoh-Mienye SA, Esenogho E, Swart TG (2020) Integrating enhanced sparse autoencoder-based artificial neural network technique and softmax regression for medical diagnosis. Electronics 9(11):1963.

[54] Mienye ID, Sun Y (2021) Improved heart disease prediction using particle swarm optimization based stacked sparse autoencoder. Electronics 10(19):2347.

[55] Aslam MA, Xue C, Chen Y, Zhang A, Liu M, Wang K, Cui D (2021) Breath analysis based early gastric cancer classification from deep stacked sparse autoencoder neural network. Scientific Reports 11(1):1-2.

[56] Gayathri JL, Abraham B, Sujarani MS, Nair MS (2022) A computer-aided diagnosis system for the classification of COVID-19 and non-COVID-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. Comput Biol Med 141:105134.

[57] Mahmoud SS, Kumar A, Tang Y, Li Y, Gu X, Fu J, Fang Q (2020) An efficient deep learning based method for speech assessment of mandarin-speaking aphasic patients. IEEE J Biomed Health Inform 24(11):3191-3202.

[58] Bastanfard A, Aghaahmadi M, Kelishami AA, Fazel M, Moghadam M (2009) Persian viseme classification for developing visual speech training application. In: Muneesawang P, Wu F, Kumazawa I, Roeksabutr A, Liao M, Tang X (eds) Advances in Multimedia Information Processing - PCM 2009. PCM 2009. Lecture Notes in Computer Science, vol 5879, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-10467-1_104.

[59] Du C, Liu C, Balamurugan P, Selvaraj P (2021) Deep learning-based mental health monitoring scheme for college students using convolutional neural network. Int J Artif Intell Tool 30(06n08):2140014.

[60] Zeberga K, Attique M, Shah B, Ali F, Jembre YZ, Chung TS (2022) A novel text mining approach for mental health prediction using Bi-LSTM and BERT model. Comput Intell Neurosci 2022.

[61] Sewani H, Kashef R (2020) An autoencoder-based deep learning classifier for efficient diagnosis of autism. Children 7(10):182.

[62] Minoofam SA, Bastanfard A, Keyvanpour MR (2022) RALF: an adaptive reinforcement learning framework for teaching dyslexic students. Multimed Tools Appl 81(5):6389-6412.

[63] Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V (2020) A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos, Solitons & Fractals140:110190.

[64] Kogilavani SV, Prabhu J, Sandhiya R, Kumar MS, Subramaniam U, Karthick A, Muhibbullah M, Imam SB (2022) COVID-19 detection based on lung CT scan using deep learning techniques. Comput Math Methods Med 2022.

[65] Xiong Z, Wang D, Liu X, Zhong F, Wan X, Li X, Li Z, Luo X, Chen K, Jiang H, Zheng M (2019) Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. J Med Chem 63(16):8749-8760.

[66] ul Qamar MT, Alqahtani SM, Alamri MA, Chen LL (2020) Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. J Pharm Anal 10(4):313-319.

[67] Ali F, El-Sappagh S, Islam SR, Ali A, Attique M, Imran M, Kwak KS (2021) An intelligent healthcare monitoring framework using

wearable sensors and social networking data. Future Gener Comput Syst 114:23-43.

[68] M Abd El-Aziz R, Alanazi R, R Shahin O, Elhadad A, Abozeid A, I Taloba A, Alshalabi R (2022) An Effective Data Science Technique for IoT-Assisted Healthcare Monitoring System with a Rapid Adoption of Cloud Computing. Comput Intell Neurosci 2022.

[69] Nasir MU, Gollapalli M, Zubair M, Saleem MA, Mehmood S, Khan MA, Mosavi A (2022) Advance genome disorder prediction model empowered with deep learning. IEEE Access 10:70317-70328.

[70] Malmir B, Amini M, Chang SI (2017) A medical decision support system for disease diagnosis under uncertainty. Expert Syst Appl 88:95-108.

[71] Khiabani SJ, Batani A, Khanmohammadi E (2022) A hybrid decision support system for heart failure diagnosis using neural networks and statistical process control. Healthc Anal 2:100110.

[72] Rajalakshmi R, Subashini R, Anjana RM, Mohan V (2018) Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye 32(6):1138-1144.

[73] Shirazi S, Baziyad H, Karimi H (2019) An Application-Based Review of Recent Advances of Data Mining in Healthcare. J Biostat Epidemiol 5(4):268-278.

[74] Fanoodi B, Malmir B, Jahantigh FF (2019) Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. Comput Biol Med 113:103415.

[75] Fukuda M, Inamoto K, Shibata N, Ariji Y, Yanashita Y, Kutsuna S, Nakata K, Katsumata A, Fujita H, Ariji E (2020) Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. Oral Radiol 36:337-343.

[76] Huang S, Yang J, Fong S, Zhao F (2020) Artificial intelligence in cancer diagnosis and prognosis. Cancer Lett 471:61–71.

[77] Wang Y, Kung L, Byrd TA (2018) Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technol Forecast Soc Change 126:3-13.

[78] Bohr A, Memarzadeh K (2020) The rise of artificial intelligence in healthcare applications. Artif Intell Healthc, pp 25-60, Academic Press.

[79] Li W, Chai Y, Khan F, Jan SR, Verma S, Menon VG, Li X (2021) A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. Mob Netw Appl 26:234-252.

[80] Alhussein M, Muhammad G (2019) Automatic voice pathology monitoring using parallel deep models for smart healthcare. IEEE Access 7:46474-46479.

[81] Mukherjee D, Mondal R, Singh PK, Sarkar R, Bhattacharjee D (2020) EnsemConvNet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. Multimed Tools Appl 79:31663-31690.

[82] Yu Z, Amin SU, Alhussein M, Lv Z (2021) Research on disease prediction based on improved DeepFM and IoMT. IEEE Access 9:39043-39054.

Enhancing Autism Severity Classification: Integrating LSTM into CNNs for Multisite Meltdown Grading

Sumbul Alam[1]*, S Pravinth Raja[2], Yonis Gulzar[3]*, Mohammad Shuaib Mir[4]
Department of Computer Science Engineering, Presidency University, Bengaluru-560064, Karnataka, India[1, 2]
Department of Management Information Systems-College of Business Administration,
King Faisal University, Al-Ahsa 31982, Saudi Arabia[3, 4]

***Abstract*—Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by deficits in social interaction, verbal and non-verbal communication, and is often associated with cognitive and neurobehavioral challenges. Timely screening and diagnosis of ASD are crucial for early educational planning, treatment, family support, and timely medical intervention. Manual diagnostic methods are time-consuming and labor-intensive, underscoring the need for automated approaches to assist caretakers and parents. While various researchers have employed machine learning and deep learning techniques for ASD diagnosis, existing models often fall short in capturing the complexity of multisite meltdowns and fully leveraging the interdependence among these meltdowns for severity assessment in acquired facial images of children, hindering the development of a comprehensive grading system. This paper introduces a novel approach using a Long Short Term Memory (LSTM) integrated Convolution Neural Network (CNN) designed to identify multisite meltdowns and exploit their interdependence for severity assessment in ASD. The process begins with image pre-processing, involving discrete convolution filters for noise removal and contrast enhancement to improve image quality. The enhanced image then undergoes instance segmentation using the Segment Anything model to identify significant regions in the child's image. The segmented region is subjected to principal component analysis for feature extraction, and these features are utilized by the LSTM-integrated CNN for meltdown detection and severity classification. The model is trained using children's images extracted from videos, and testing is performed on videos captured during children's observations. Performance analysis reveals superior results, with a training accuracy of 88% and validation accuracy of 84%, outperforming conventional methods. This innovative approach not only enhances the efficiency of ASD diagnosis but also provides a more nuanced understanding of multisite meltdowns and their impact on severity, contributing to the development of a robust grading system.**

Keywords*—Autism spectrum disorder; mutilating meltdown; convolution neural network; long short term memory; multisite meltdown; video classification; image classification

I. Introduction

Autism spectrum disorder (ASD) represents a complex neurodevelopmental syndrome characterized by a wide range of challenges in verbal and nonverbal communication skills, as well as behavioral and social interactions [1]. While ASD can manifest at any age, it typically becomes evident around the age of 2 or 3 when children start to withdraw, exhibit distinct behaviors, and present challenges in social engagement. The etiology of this disorder is diverse, and the underlying neurodevelopmental mechanisms are not fully understood [2].

The detection of a high degree of autism severity in a child is particularly concerning, as it often leads to the development of more frequent meltdowns. These meltdowns not only pose a risk of self-injury to the child but can also result in harm to caregivers or parents [3]. Early diagnosis of ASD is crucial, offering significant benefits in terms of intellectual development, adaptive behavior, and the reduction of overall severity. The advent of noninvasive acquisition technology has made disease diagnosis more feasible, but manual diagnosis remains a highly challenging and labor-intensive task. Consequently, there is a pressing need to develop an automated disease diagnosis tool for ASD.

In recent times, the pervasive influence of artificial intelligence (AI) has become increasingly apparent, bringing about transformative changes across a spectrum of fields and enriching various facets of our everyday existence [4,5]. It has redefined how we approach education [6], fine-tuned financial strategies [7], simplified agricultural workflows [8–16], and elevated healthcare diagnostics to new heights [17–23]. As it seamlessly integrates into these diverse sectors, AI continues to demonstrate its capacity for generating unparalleled efficiencies, refining decision-making procedures, and addressing intricate challenges with a precision derived from data-driven insights[24,25].

Researchers have turned to machine learning and deep learning approaches to enhance the accuracy of ASD diagnosis. These models leverage the concept of correlation and co-variation among spatio-temporal data. Specifically, Convolutional Neural Networks (CNNs) have proven highly capable of extracting features based on spatio-temporal descriptors to classify various gestures exhibited by ASD children. However, existing models have limitations, as they fail to compute the appearance of multisite meltdowns and fully exploit the interdependence among these meltdowns for severity computation in the acquired facial images of children, hindering the development of a comprehensive grading system [26].

This paper presents a novel approach, introducing a Long Short-Term Memory (LSTM) integrated Convolutional Neural Network designed to detect multisite meltdowns and exploit their interdependence for severity computation in ASD. The proposed methodology involves initial image pre-processing

*Correspondence: sumbulalam@presidencyuniversity.in (S.A); ygulzar@kfu.edu.sa (Y.G.)

using contrast enhancement and histogram equalization to improve image quality. The enhanced image is then subjected to an instance segmentation technique, termed the "segment anything" model, to identify significant regions in the child's image. These segmented regions undergo principal component analysis for feature extraction, which is then employed in the novel LSTM-integrated CNN [27] for meltdown detection and severity classification, achieving increased accuracy. The model is trained using images of children acquired from videos, and testing is conducted using videos acquired during children's observations.

The subsequent sections of the article are organized as follows: Section II provides a detailed problem statement and a review of the literature on detecting the meltdown state of autism spectrum disorder. Section III outlines the proposed deep learning methodology for detecting and classifying multisite meltdown states and their severity to establish a grading system. Section IV presents an experimental analysis of the proposed methodology on the disease dataset, including performance metrics such as accuracy through a confusion matrix. Finally, Section V concludes the work and offers future suggestions.

II. Related Work

In this part, numerous conventional approaches using machine learning and deep learning architecture to detect the autism spectrum disorder along meltdown state using behavioral and kinematics features has been detailed as follows:

A. *Autism Spectrum Disorder Detection using Restricted Kinematic Features*

In this literature, machine learning model is used to detect the Autism spectrum disorder using the restricted kinematic features from the kinematic data. In this computed on basis of movement during motor task. Entropy, amplitude, velocity, acceleration were considered as kinematic features and considered as symptom of the ASD. Machine learning classifier such as support vector machine were employed to classify or detect the ASD using the feature values which has cognitive flexibility and it has high manifestation [28].

B. *Autism Spectrum Disorder Detection using Autoencoder-based Support Vector Machine - Recursive Feature Elimination Technique*

In this literature, deep learning architecture is used to detect the ASD through functional connectivity features of the multiple regions. Functional connectivity feature is extracted and those features are employed to Recursive feature elimination technique to select the primitive features. Next, Autoencoder model is used to extract the high latent features and complicated features. It is considered as optimal features [29]. Those features were employed to softmax classifier which employs the Support vector machine to detect the ASD.

C. *Autism Spectrum Disorder and Meltdown Detection on Facial Geometric Features using Recurrent Neural Network*

In this literature, recurrent neural network is employed to detect and classify the autism spectrum disorder during the meltdown crisis. Initially model extracts the micro facial expression of children as geometric features and detects the child with autism or without autism. On detection of autism state , children is classified with severity of meltdown Hidden layer of the model process the feature to produce the optimal severity state of the children with meltdown [30].

D. *Autism Spectrum Disorder and Meltdown Detection using Recurrent Attention Network on Morphological Features*

In this literature, recurrent attention network is employed to detect and classify the autism spectrum disorder during the meltdown crisis. Initially model extracts the morphological features and detects the child with autism or without autism. On detection of autism state, children are classified with severity of meltdown. Attention layer of the model process the feature embedding to produce the optimal severity state of the children with meltdown with high efficiency [31].

E. *Autism Spectrum Disorder and Meltdown Detection using Deep Neural Network on Audio-based Features*

In this literature, deep neural network is employed to detect and classify the autism spectrum disorder during the meltdown crisis using audio based signals. Initially model extract the audio based features on transforming the speech data to mel spectrogram and those audio based feature like pitch, RMS and MFFS used to detect the child with autism or without autism. On detection of autism state, children are classified with severity of meltdown. Dense layer of the model process the features as embedding of features to produce the optimal severity state of the children with meltdown with high efficiency [32].

III. Proposed Model

In this section, we introduce a sophisticated system for grading multisite meltdowns and classifying the severity of autism spectrum disorder in children. Our approach integrates a Long Short-Term Memory (LSTM) with a Convolutional Neural Network (CNN) specifically tailored for assessing the intensity of meltdowns across various sites, encompassing expressions of distress such as crying, screaming, and stimming. The step-by-step processing to achieve this overarching objective is detailed below:

A. *Image Preprocessing- Discrete Convolution Filter*

This section outlines the application of noise reduction and contrast enhancement techniques to the acquired training images through the utilization of a discrete convolution filter. The initial phase of noise reduction in training images serves to eliminate blurriness, while the subsequent contrast enhancement enhances the overall image quality. The discrete convolution filter method, employed in this process, effectively heightens the sharpness of object edges within each image [33].

The discrete convolution filter method is instrumental in achieving this enhancement, producing a pronounced increase in edge sharpness. A special case of this method involves the averaging of brightness values, exemplified by the formula:

$$\text{f(i,j)=w*h=}\sum_{m=-a}^{a}\sum_{n=-b}^{b} w(m,n)h(i+m,j+n) \quad (1)$$

This Formula depicts a linear operation wherein the resulting value in the output image pixel f(i,j) is calculated as a linear combination of the brightness in a local neighborhood of the pixel h(i,j) in the input image. The function (w) in this context represents the convolution kernel, encapsulating the linear operations involved in this discrete convolution process.

B. Instance Segmentation - Segment Anything Model

The enhanced image undergoes application of the instance segmentation technique, known as the "Segment Anything" model, aimed at delineating significant regions within the child image. This model excels in segmenting every pixel with similar values, employing boundary probability algorithms (gPb and UCM) to calculate edge and its weight. A threshold is then applied for each pixel and its adjacent pixel to refine the segmentation process [34].

The instance segmentation process, as expressed by the following formula:

$$L(x,y)= \sum_{i=0}^{i<x,} \sum_{j=0}^{y<j} I(i,j) \text{ where } 0<x<N \text{ and } 0<y<M \quad (2)$$

Where $0<x<N$ and $0<y<M$, involves aggregating pixel values within the specified range, providing a comprehensive representation of the preprocessed image. To integrate similar pixels within edges, a connected component approach is employed. The result is a hierarchical organization of regions obtained through a coarse-to-fine process.

C. Feature Extraction - Principle Component Analysis

The application of Principal Component Analysis (PCA) serves as a pivotal step in discerning the normal and meltdown states of children with autism. PCA, leveraging spatio-temporal information within each frame, identifies highly discriminating features. The execution of PCA results in a set of training data characterized by disconnectedness among data points and dense similarity within classes [35].

The analysis extends to examining segmented objects across two consecutive frames of images, calculating and defining features at various focal points such as the left eye, right eye, right mouth corner, left mouth corner, and nose. Each principal component represents the maximum variance among these focal points. To handle the complexity of computing features in high dimensions, PCA effectively minimizes dimensions without significant loss of feature information through matrix formation and distance calculation.

The composed feature vectors encapsulate facial interest components. Given an image of size $N×N$, it is initially transformed into a 1D vector U, housing variance values of substantial magnitude. The variance for a specified feature X in an image is calculated by the formula:

$$\text{variance(y)} = \frac{\sum_{i=1}^{n} \text{b(yi – y) (yi – y)}}{n-1} \quad (3)$$

Furthermore, the covariance of features is calculated for the objects X and Y that change together with the mean, expressed as:

$$\text{Covariance(y,x)} = \frac{\sum_{i=1}^{n} \text{a(yi – y) (xi – x)}}{n-1} \quad (4)$$

The resulting Covariance Matrix, a $N×N$ feature matrix, is represented by:

$$M_{ij}\text{= Covariance(x,y)} \quad (5)$$

D. Eigen Vector Analysis for Facial Feature Classification

The computation involves deriving the Eigen vector, denoted as M_{ij}^{Eigen} , serving as a feature vector that encapsulates principle feature groups. These feature values are accompanied by Eigen values and are pivotal for the subsequent classification of facial features.

For each meltdown and normal state, the associated feature values fall within the range of 0 to 1. Here, 0 signifies the normal state, while 1 designates the meltdown state. Table I presents the Eigen vector, composed of features extracted specifically for the meltdown state.

TABLE I. LIST OF FEATURE EXTRACTED FOR MELTDOWN STATE

Feature	Description
Eyes Closed	Distance between two eye lids
Mouth Open	Distance between two lips
Lips enlargement	Radius of the lips
Object in ears	Hand in the ears
Object in head	Hand in the head or hair

These features provide a detailed description of facial expressions during the meltdown state, including eye and mouth behaviors, lip enlargement, and hand placements indicative of distress. The Eigen vector, with its associated Eigen values, serves as a valuable tool for the effective classification of these facial features, contributing to a comprehensive understanding of autism states.

E. Long Short Term Memory Integrated Convolution Neural Network

The extracted features play a pivotal role in the functionality of the novel Long Short-Term Memory Integrated Convolutional Neural Network (LSTM-CNN) designed for the detection and classification of meltdown states, with a focus on assessing the severity. This innovative approach leverages hyperparameter-optimized layers to enhance the processing of information [36].

1) Long short term memory: The Long Short-Term Memory (LSTM) component within our integrated model is a key element endowed with the unique ability to learn and comprehend long-term dependencies through its intricate network connections. It excels in the storage of pertinent information within a memory cell while effectively discarding extraneous details. Each LSTM unit comprises a memory cell equipped with two gates: input and output, and a forget gate. In the context of our research, the LSTM model serves as a repository for multisite meltdown feature maps, a product of the convolution and max-pooling layers in the Convolutional Neural Network (CNN). These feature maps encapsulate crucial patterns essential for the accurate detection and classification of meltdown states.

In order to fine-tune the performance of our LSTM-CNN model, we employ hyperparameter optimization, as outlined in Table II.

TABLE II. HYPERPARAMETER TUNING

Hyperparameter	Value
Learning rate	10^{-6}
Epoch Value	100
Activation function	ReLu
Loss Function	Cross Entropy

The hyperparameters, meticulously chosen and specified in Table II, play a pivotal role in shaping the learning dynamics of our model. The learning rate, set at 10^-6, determines the step size during optimization, ensuring a balance between accuracy and efficiency. The epoch value of 100 signifies the number of times the entire dataset is processed during training, influencing the model's convergence. The ReLU activation function is employed to introduce non-linearity, enhancing the model's capacity to learn intricate patterns. Finally, the Cross Entropy loss function measures the dissimilarity between predicted and actual values, guiding the model towards optimal performance. This comprehensive hyperparameter tuning aims to maximize the LSTM-CNN model's efficacy in detecting and classifying meltdown states with a nuanced understanding of their severity.

2) Convolution neural network: The Convolutional Neural Network (CNN) serves as a cornerstone in our methodology, tasked with processing the intricately extracted features across multiple layers. Its primary objective is to adeptly detect and classify the multisite meltdown state, discerning the severity level inherent in each case. The CNN's architecture is meticulously designed, incorporating essential elements such as the convolution layer, max-pooling layer, and fully connected layer.

The convolution layer plays a critical role in feature extraction, employing filters to scan and identify distinctive patterns within the input data. This process enables the CNN to capture essential spatial hierarchies and dependencies in the multisite meltdown features. Subsequently, the max-pooling layer strategically downsizes the spatial dimensions of the extracted features, promoting computational efficiency and reducing the risk of overfitting.

The fully connected layer, a crucial component of the CNN architecture, is responsible for processing linear features extracted from the preceding layers. It incorporates an activation function to introduce non-linearity, allowing the model to learn complex relationships within the data. Furthermore, the softmax function within the fully connected layer serves a dual purpose – detection and classification. It assigns probabilities to different meltdown states, facilitating a nuanced understanding of the severity levels associated with each classification.

To guide the training process effectively, a loss function is integrated into the fully connected layer. This function calculates the classification error, providing feedback to the model during the training phase. The objective is to minimize this error, enhancing the CNN's ability to accurately detect and classify multisite meltdown states.

For a visual representation of our proposed model's architecture, refer to Fig. 1. This diagram encapsulates the intricate interplay between the convolution layer, max-pooling layer, and fully connected layer, offering a comprehensive overview of the network's structure and functionality. The synergy of these components within the CNN underscores its efficacy in robustly addressing the detection and classification challenges posed by multisite meltdown scenarios.

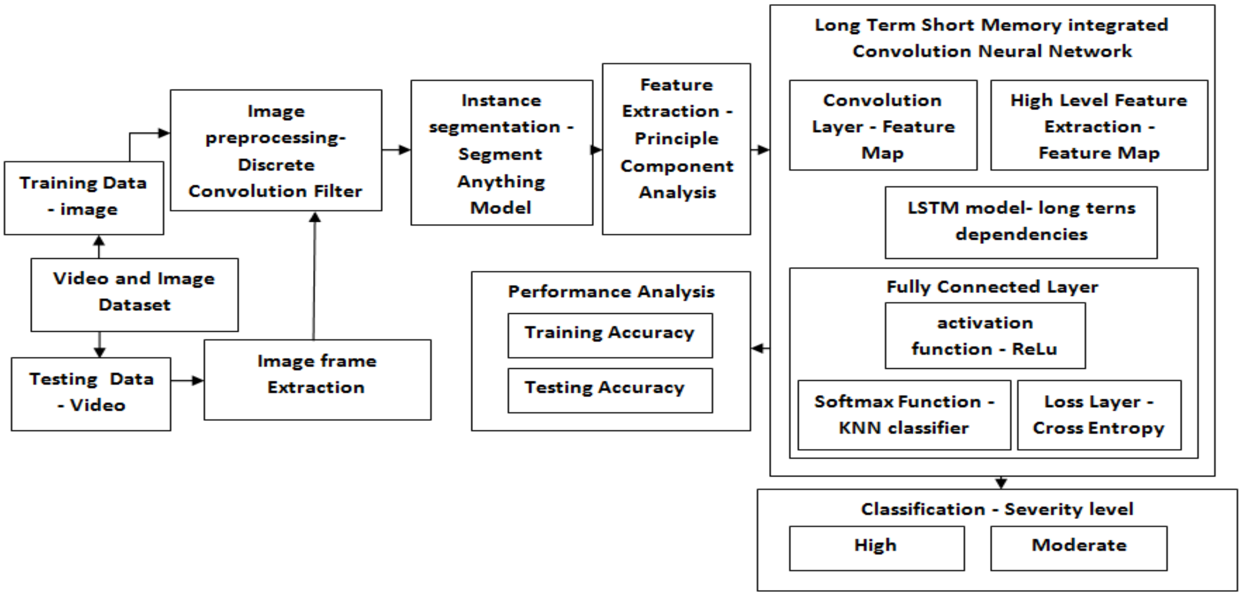

Training Data - image
Video and Image Dataset
Testing Data - Video
Image preprocessing- Discrete Convolution Filter
Image frame Extraction
Instance segmentation - Segment Anything Model
Feature Extraction - Principle Component Analysis
Long Term Short Memory integrated Convolution Neural Network
Convolution Layer - Feature Map
High Level Feature Extraction - Feature Map
LSTM model- long terns dependencies
Fully Connected Layer
activation function - ReLu
Softmax Function - KNN classifier
Loss Layer - Cross Entropy
Performance Analysis
Training Accuracy
Testing Accuracy
Classification - Severity level
High
Moderate

Fig. 1. Proposed architecture.

- Convolution Layer: The Convolution Layer is a pivotal element in our architecture, containing multiple filters and kernels that convolve with features extracted from different regions. This convolution operation produces a feature map that represents the underlying patterns in the meltdown state. Mathematically, convolution involves the multiplication of the feature vector, containing information about the meltdown state in a specified region, with multiple filters [37]. The convolution process is expressed as in the Formula:

$$x_n = \sum_{K=0}^{N-1} Y_K F_{n-k} \tag{6}$$

Here, *Y* represents the feature, and *F* is the filter.

The convolution layer generates a feature map through convolution operations, encompassing both low-level and latent features. The convergence of this feature map is facilitated by epochs, incrementally increasing feature generation. Normalization through the Rectified Linear Unit (ReLU) activation function further refines the feature map, obtaining a linear representation. The cosine distance measure is then employed to compute the distance among features.

- Pooling Layer: The Pooling Layer follows the convolutional operations, serving to further reduce the features obtained from the convolution layer. This step is crucial for high-level meltdown feature extraction and is essentially a form of down-sampling, diminishing the dimensions of facial features and retaining only selected weighted meltdown features. The Max Pooling layer plays a vital role in connecting the meltdown features into small patches, estimating the maximum number of features for each subset. This process enhances model generalization [38].

- Long Short Term Memory Layer: The LSTM Layer is employed to store the feature map derived from the convolution and max-pooling layers of the Convolutional Neural Network. It excels in preserving long-term dependencies through its intricate network connections. The stored meltdown features, each assigned a weight value, reside in the memory cell of each LSTM unit. These features, converted into a feature matrix, are input into the LSTM for the fusion of multisite meltdown information [39].

$$C_t = \tanh(X_t * V_t + H_{t-1} * W_t) \tag{7}$$

In the CNN-LSTM hybrid model, normal and meltdown features are extracted from both the convolutional and LSTM layers. The ordering of these features is then utilized, where *Ht* represents cell memory information, and *Wt* represents the weight vector.

- Dense Layer – Fully Connected Layer: The Dense Layer, organized as a fully connected layer, processes the feature map composed of multisite meltdown features across facial regions. This layer extracts discriminative features related to crying, stimming, and screaming. The activation function is applied for feature normalization and flattening, addressing non-linearity and overfitting concerns in the feature maps.

- Softmax Layer: The Softmax Layer, integrated into the fully connected layer, is crucial for detecting the meltdown state. It combines these states, assigning aggregate weights to identify the severity of multisite meltdown states using a Naive Bayes classifier. A loss layer is further incorporated to minimize feature variance across classes. The Softmax function, as represented in Formula (8), calculates the probability distribution:

$$\text{Softmax Function } P_j = \frac{e^{x_j}}{\sum_1^k e^{x_k}} \tag{8}$$

where, e^{xj} is the feature map long dependency vector.

- Classifier and Decision Rule: The feature vector is projected for classification by applying Bayes theorem, utilizing a maximum likelihood function to aggregate similar emotion features of autistic children based on feature values. The final classification result is generated by integrating results according to the decision rule. Feature values related to crying, screaming, and stimming form distinct classes. The Maximum Likelihood function, as given in Eq.9, incorporates the class, feature values, and density function:

$$f_n(y, \theta) = \prod_{k=1}^{n} f_k \tag{9}$$

Here, *Y* is the class of the meltdown, θ is the vector containing feature values, and f_k is the density function. This comprehensive approach ensures a robust and nuanced classification of multisite meltdown states based on their severity levels.

Algorithm 1: Multisite Meltdown Detection and Severity Classification.

```
Input : Video and Images of the Child observations
Output: Detection of Multisite Meltdown and severity classes
Process
Train()
Preprocessing of training images ()
Contrast Enhancement ()
Preprocessed image =Discrete Convolution filter (Training images)
Instance Segmentation()
Segment = Segment Anything Model( Preprocessed image )
Feature extraction()
Transform the image pixel of segment into matrix
Compute the covariance and correlation on the matrix
Determine the eigen value and eigen vector
Eigen vector = Feature Vector
Disease detection and classification ()
Convolution Neural Network()
Convolution Layer () = VGG19()
Low level feature  = Kernel (Feature vector)
Feature map = ReLu( Low level Features )
            Max pooling layer ()
High level feature = Kernel (Feature vector)
Feature map = ReLu( Low level Features )
            LTSM layer ()
```

```
        Store feature map as long terms dependencies
        Feature Dependencies=  combining the multiple meltdown
        Fully connected layer ()
        Activation function = ReLu()
        Softmax function = Naive Bayes ( Feature dependencies map)
        Detection of disease = { Crying , Screaming , Stimming}
        Severity Class= { High , Moderate }
        Loss function - Cross Entropy()
```

The algorithm offers a holistic approach to Multisite Meltdown Detection and Severity Classification. It integrates preprocessing, segmentation, feature extraction, and a robust combination of Convolution Neural Network and Long Short-Term Memory for accurate and nuanced results. The inclusion of a variety of techniques, such as contrast enhancement, instance segmentation, and feature mapping, contributes to the algorithm's effectiveness in handling complex scenarios related to autism severity classification. The architecture demonstrates a deep understanding of both low-level and high-level features, providing a comprehensive solution for the challenging task at hand.

IV. Experimental Results

In this pivotal section, we delve into a thorough analysis of the experimental outcomes, leveraging cross-fold validation on a simulated dataset within the Python environment [40]. The performance evaluation of our proposed architecture for autism severity classification is meticulously conducted, with optimal parameters defining the model's performance. The implementation utilizes the versatile Scikit-learn package, incorporating various machine learning algorithms, and OpenCV for efficient image processing and preparation.

A. Dataset Description - Meltdown Crisis

The cornerstone of our investigation lies in the Meltdown Crisis dataset, a robust collection designed for identifying multisite meltdown severity in autism children. Comprising 59 videos, this dataset offers a detailed narrative, encompassing facial expressions and physical activities during the meltdown crisis. The dataset is structured into emotion frames and non-emotion frames, categorizing various states such as normal, post-crisis, and meltdown crisis states. For streamlined evaluation, the dataset is strategically partitioned into training, testing, and validation sets [41].

The training phase involves the utilization of images extracted from children in the videos, while testing is conducted on videos observed during children's activities. Fig. 2 presents the confusion matrix for the validation dataset, consisting of 59 videos. This visual representation encapsulates the model's proficiency in classifying instances across different categories.

Fig. 3 provides a comprehensive snapshot of the training and validation accuracy of our model. This graphical representation elucidates the learning trajectory of the model over multiple epochs. The model achieves commendable results, boasting an 88% training accuracy and an 84% validation accuracy, reflecting its robust learning capabilities.

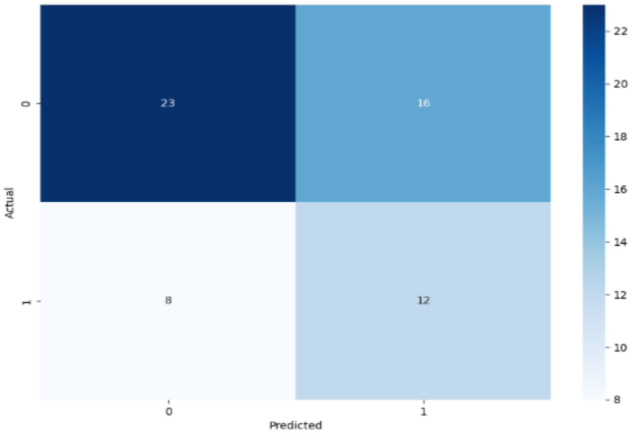


Fig. 2. Confusion matrix.

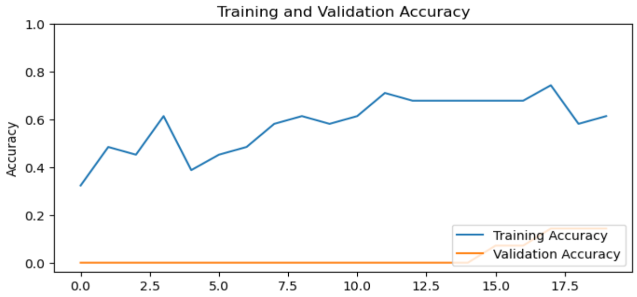


Fig. 3. Training and validation accuracy of the model.

Performance analysis extends to the examination of the training and validation loss, as illustrated in Fig. 4. These curves offer insights into the model's optimization process, demonstrating a balanced and decreasing trend in error reduction over the course of training and validation.

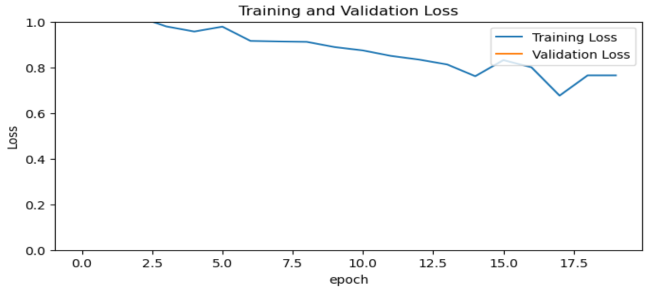


Fig. 4. Training and validation of the model.

TABLE III. Performance Evaluation of the Model

Technique	Accuracy		Loss	
	Training	Validation	Training	Validation
CNN+LSTM	88	84	0.8	0.1

Table III meticulously encapsulates the quantitative results of our model. The training accuracy, validation accuracy, training loss, and validation loss are presented, offering a detailed performance snapshot. The CNN–LSTM models were meticulously trained for 100 epochs, employing a batch size of 128, and utilizing a cross-entropy loss function. These parameters were strategically chosen to ensure a robust and effective training process.

The proposed model exhibits superior performance when benchmarked against conventional approaches. With a training accuracy of 88% and a validation accuracy of 84%, coupled with minimal training and validation loss, our CNN–LSTM model demonstrates efficacy in autism severity classification. These findings underscore the potential of our integrated architecture in providing nuanced and accurate assessments in the context of multisite meltdown grading.

V. Conclusion

In this study, we introduced a novel approach for detecting multisite meltdowns in children with Autism Spectrum Disorder (ASD) using a Long Short-Term Memory (LSTM) [42] integrated Convolutional Neural Network (CNN). Our designed architecture aims to leverage the dependency among multisite meltdowns to enhance the severity computation, ultimately contributing to the development of a robust grading system. The comprehensive pipeline of our model encompasses pre-processing techniques to enhance image quality, the utilization of the Segment Anything model for segmentation, and principle component analysis for feature extraction. These steps are crucial in isolating significant regions within child images and extracting pertinent features.

The extracted features are subsequently fed into the CNN+LSTM classifier, which effectively detects and classifies multisite meltdowns, providing valuable insights into their severity. The model's performance analysis yielded promising results, with a training accuracy of 88% and a validation accuracy of 84%. This underscores the efficacy of our proposed architecture in accurately identifying and grading multisite meltdowns in children with ASD.

While our current model has shown promising results, there are avenues for future research and improvement. Firstly, the inclusion of a larger and more diverse dataset could enhance the model's generalization capabilities. Exploring advanced techniques for feature extraction and segmentation may further refine the model's ability to capture subtle nuances in facial expressions during meltdowns.

Additionally, investigating real-time applications and deployment in clinical settings could provide valuable insights into the model's practical utility. Fine-tuning hyperparameters and exploring alternative neural network architectures may also contribute to optimizing the model's performance.

Acknowledgment

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under the Project GRANT5, 306.

References

[1] Alam, S.; Raja, P.; Gulzar, Y. Investigation of Machine Learning Methods for Early Prediction of Neurodevelopmental Disorders in Children. Wirel Commun Mob Comput 2022, 2022.

[2] Guha, T.; Yang, Z.; Ramakrishna, A.; Grossman, R.B.; Darren, H.; Lee, S.; Narayanan, S.S. On Quantifying Facial Expression-Related Atypicality of Children with Autism Spectrum Disorder. Proc IEEE Int Conf Acoust Speech Signal Process 2015, 2015, 803–807, doi:10.1109/ICASSP.2015.7178080.

[3] Guo, J.; Zhou, S.; Wu, J.; Wan, J.; Zhu, X.; Lei, Z.; Li, S.Z. Multi-Modality Network with Visual and Geometrical Information for Micro Emotion Recognition. ieeexplore.ieee.orgJ Guo, S Zhou, J Wu, J Wan, X Zhu, Z Lei, SZ Li2017 12th IEEE international conference on automatic face, 2017•ieeexplore.ieee.org.

[4] Ayoub, S.; Gulzar, Y.; Reegu, F.A.; Turaev, S. Generating Image Captions Using Bahdanau Attention Mechanism and Transfer Learning. Symmetry (Basel) 2022, 14, 2681.

[5] Hamid, Y.; Elyassami, S.; Gulzar, Y.; Balasaraswathi, V.R.; Habuza, T.; Wani, S. An Improvised CNN Model for Fake Image Detection. International Journal of Information Technology 2023, 15, 5–15, doi:10.1007/S41870-022-01130-5.

[6] Sahlan, F.; Hamidi, F.; Misrat, M.Z.; Adli, M.H.; Wani, S.; Gulzar, Y. Prediction of Mental Health Among University Students. International Journal on Perceptive and Cognitive Computing 2021, 7, 85–91.

[7] Gulzar, Y.; Alwan, A.A.; Abdullah, R.M.; Abualkishik, A.Z.; Oumrani, M. OCA: Ordered Clustering-Based Algorithm for E-Commerce Recommendation System. Sustainability 2023, Vol. 15, Page 2947 2023, 15, 2947, doi:10.3390/SU15042947.

[8] Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. Sustainability 2023, 15, 1906.

[9] Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. Sustainability 2023, 15, 901.

[10] Dhiman, P.; Kaur, A.; Balasaraswathi, V.R.; Gulzar, Y.; Alwan, A.A.; Hamid, Y. Image Acquisition, Preprocessing and Classification of Citrus Fruit Diseases: A Systematic Literature Review. Sustainability 2023, Vol. 15, Page 9643 2023, 15, 9643, doi:10.3390/SU15129643.

[11] Albarrak, K.; Gulzar, Y.; Hamid, Y.; Mehmood, A.; Soomro, A.B. A Deep Learning-Based Model for Date Fruit Classification. Sustainability 2022, 14.

[12] Hamid, Y.; Wani, S.; Soomro, A.B.; Alwan, A.A.; Gulzar, Y. Smart Seed Classification System Based on MobileNetV2 Architecture. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT); IEEE, 2022; pp. 217–222.

[13] Gulzar, Y.; Hamid, Y.; Soomro, A.B.; Alwan, A.A.; Journaux, L. A Convolution Neural Network-Based Seed Classification System. Symmetry (Basel) 2020, 12, 2018.

[14] Gulzar, Y.; Ünal, Z.; Akta¸s, H.A.; Mir, M.S. Harnessing the Power of Transfer Learning in Sunflower Disease Detection: A Comparative Study. Agriculture 2023, Vol. 13, Page 1479 2023, 13, 1479, doi:10.3390/AGRICULTURE13081479.

[15] Malik, I.; Ahmed, M.; Gulzar, Y.; Baba, S.H.; Mir, M.S.; Soomro, A.B.; Sultan, A.; Elwasila, O. Estimation of the Extent of the Vulnerability of Agriculture to Climate Change Using Analytical and Deep-Learning Methods: A Case Study in Jammu, Kashmir, and Ladakh. Sustainability 2023, Vol. 15, Page 11465 2023, 15, 11465, doi:10.3390/SU151411465.

[16] Aggarwal, S.; Gupta, S.; Gupta, D.; Gulzar, Y.; Juneja, S.; Alwan, A.A.; Nauman, A. An Artificial Intelligence-Based Stacked Ensemble Approach for Prediction of Protein Subcellular Localization in Confocal Microscopy Images. Sustainability 2023, Vol. 15, Page 1695 2023, 15, 1695, doi:10.3390/SU15021695.

[17] Gulzar, Y.; Khan, S.A. Skin Lesion Segmentation Based on Vision Transformers and Convolutional Neural Networks—A Comparative Study. Applied Sciences 2022, Vol. 12, Page 5990 2022, 12, 5990, doi:10.3390/APP12125990.

[18] Mehmood, A.; Gulzar, Y.; Ilyas, Q.M.; Jabbari, A.; Ahmad, M.; Iqbal, S. SBXception: A Shallower and Broader Xception Architecture for Efficient Classification of Skin Lesions. Cancers 2023, Vol. 15, Page 3604 2023, 15, 3604, doi:10.3390/CANCERS15143604.

[19] Khan, F.; Ayoub, S.; Gulzar, Y.; Majid, M.; Reegu, F.A.; Mir, M.S.; Soomro, A.B.; Elwasila, O. MRI-Based Effective Ensemble Frameworks for Predicting Human Brain Tumor. Journal of Imaging 2023, Vol. 9, Page 163 2023, 9, 163, doi:10.3390/JIMAGING9080163.

[20] Majid, M.; Gulzar, Y.; Ayoub, S.; Khan, F.; Reegu, F.A.; Mir, M.S.; Jaziri, W.; Soomro, A.B. Enhanced Transfer Learning Strategies for Effective Kidney Tumor Classification with CT Imaging. International

Journal of Advanced Computer Science and Applications 2023, 14, 2023, doi:10.14569/IJACSA.2023.0140847.

[21] Anand, V.; Gupta, S.; Gupta, D.; Gulzar, Y.; Xin, Q.; Juneja, S.; Shah, A.; Shaikh, A. Weighted Average Ensemble Deep Learning Model for Stratification of Brain Tumor in MRI Images. Diagnostics 2023, Vol. 13, Page 1320 2023, 13, 1320, doi:10.3390/DIAGNOSTICS13071320.

[22] Majid, M.; Gulzar, Y.; Ayoub, S.; Khan, F.; Reegu, F.A.; Mir, M.S.; Jaziri, W.; Soomro, A.B. Using Ensemble Learning and Advanced Data Mining Techniques to Improve the Diagnosis of Chronic Kidney Disease. International Journal of Advanced Computer Science and Applications 2023, 14, doi:10.14569/IJACSA.2023.0141050.

[23] Khan, S.A.; Gulzar, Y.; Turaev, S.; Peng, Y.S. A Modified HSIFT Descriptor for Medical Image Classification of Anatomy Objects. Symmetry (Basel) 2021, 13, 1987.

[24] Dhiman, P.; Bonkra, A.; Kaur, A.; Gulzar, Y.; Hamid, Y.; Mir, M.S.; Soomro, A.B.; Elwasila, O. Healthcare Trust Evolution with Explainable Artificial Intelligence: Bibliometric Analysis. Information 2023, Vol. 14, Page 541 2023, 14, 541, doi:10.3390/INFO14100541.

[25] Hanafi, M.F.F.M.; Nasir, M.S.F.M.; Wani, S.; Abdulghafor, R.A.A.; Gulzar, Y.; Hamid, Y. A Real Time Deep Learning Based Driver Monitoring System. International Journal on Perceptive and Cognitive Computing 2021, 7, 79–84.

[26] Masmoudi, M.; … S.J.-2019 15th I.; 2019, undefined Meltdowncrisis: Dataset of Autistic Children during Meltdown Crisis. ieeexplore.ieee.orgM Masmoudi, SK Jarraya, M Hammami2019 15th International Conference on Signal-Image Technology, 2019•ieeexplore.ieee.org.

[27] Haweel, R.; Dekhil, O.; Shalaby, A.; Mahmoud, A.; Ghazal, M.; Khalil, A.; Keynton, R.; Barnes, G.; El-Baz, A. A Novel Framework for Grading Autism Severity Using Task-Based FMRI. Proceedings - International Symposium on Biomedical Imaging 2020, 2020-April, 1404–1407, doi:10.1109/ISBI45749.2020.9098430.

[28] Herringshaw, A.J.; Ammons, C.J.; DeRamus, T.P.; Kana, R.K. Hemispheric Differences in Language Processing in Autism Spectrum Disorders: A Meta-Analysis of Neuroimaging Studies. Autism Research 2016, 9, 1046–1057, doi:10.1002/AUR.1599.

[29] Redcay, E.; Courchesne, E. Deviant Functional Magnetic Resonance Imaging Patterns of Brain Activity to Speech in 2-3-Year-Old Children with Autism Spectrum Disorder. Biol Psychiatry 2008, 64, 589–598, doi:10.1016/j.biopsych.2008.05.020.

[30] Jarraya, S.K.; Masmoudi, M.; Hammami, M. Compound Emotion Recognition of Autistic Children During Meltdown Crisis Based on Deep Spatio-Temporal Analysis of Facial Geometric Features. IEEE Access 2020, 8, 69311–69326, doi:10.1109/access.2020.2986654.

[31] Ke, F.; Yang, R. Classification and Biomarker Exploration of Autism Spectrum Disorders Based on Recurrent Attention Model. IEEE Access 2020, 8, 216298–216307, doi:10.1109/access.2020.3038479.

[32] Eni, M.; Dinstein, I.; Ilan, M.; Menashe, I.; Meiri, G.; Zigel, Y. Estimating Autism Severity in Young Children From Speech Signals Using a Deep Neural Network. IEEE Access 2020, 8, 139489–139500, doi:10.1109/access.2020.3012532.

[33] Liu, W.; Li, M.; Yi, L. Identifying Children with Autism Spectrum Disorder Based on Their Face Processing Abnormality: A Machine Learning Framework. Autism Research 2016, 9, 888–898, doi:10.1002/aur.1615.

[34] Rudovic, O.; Utsumi, Y.; Lee, J.; Hernandez, J.; Ferrer, E.C.; Schuller, B.; Picard, R.W. CultureNet: A Deep Learning Approach for Engagement Intensity Estimation from Face Images of Children with Autism. IEEE International Conference on Intelligent Robots and Systems 2018, 339–346, doi:10.1109/IROS.2018.8594177.

[35] Lombardo, M. V; Pramparo, T.; Gazestani, V.; Warrier, V.; Bethlehem, R.A.I.; Carter Barnes, C.; Lopez, L.; Lewis, N.E.; Eyler, L.; Pierce, K.; et al. Large-Scale Associations between the Leukocyte Transcriptome and BOLD Responses to Speech Differ in Autism Early Language Outcome Subtypes. Nat Neurosci 2018, 21, 1680–1688, doi:10.1038/s41593-018-0281-3.

[36] Lord, C.; Elsabbagh, M.; Baird, G.; Veenstra-Vanderweele, J. Autism Spectrum Disorder. Lancet 2018, 392, 508–520, doi:10.1016/S0140-6736(18)31129-2.

[37] Gotham, K.; Pickles, A.; Lord, C. Standardizing ADOS Scores for a Measure of Severity in Autism Spectrum Disorders. J Autism Dev Disord 2009, 39, 693–705, doi:10.1007/s10803-008-0674-3.

[38] Stoner, R.; Chow, M.L.; Boyle, M.P.; Sunkin, S.M.; Mouton, P.R.; Roy, S.; Wynshaw-Boris, A.; Colamarino, S.A.; Lein, E.S.; Courchesne, E. Patches of Disorganization in the Neocortex of Children with Autism. N Engl J Med 2014, 370, 1209–1219, doi:10.1056/NEJMoa1307491.

[39] Karten, A.; Hirsch, J. Brief Report: Anomalous Neural Deactivations and Functional Connectivity during Receptive Language in Autism Spectrum Disorder: A Functional MRI Study. J Autism Dev Disord 2015, 45, 1905–1914, doi:10.1007/s10803-014-2344-y.

[40] Kliuev, E.A.; Sheyko, G.E.; Dunayev, M.G.; Abramov, S.A.; Dvoryaninova, V. V; Balandina, O. V; Karyakin, N.N.; Belova, A.N. The Role of Functional MRI in Understanding the Origin of Speech Delay in Autism Spectrum Disorders. Sovremennye tehnologii v medicine 2019, 11, 66, doi:10.17691/stm2019.11.3.09.

[41] Ayoub, S.; Gulzar, Y.; Rustamov, J.; Jabbari, A.; Reegu, F.A.; Turaev, S. Adversarial Approaches to Tackle Imbalanced Data in Machine Learning. Sustainability 2023, Vol. 15, Page 7097 2023, 15, 7097, doi:10.3390/SU15097097.

[42] Khan, F.; Gulzar, Y.; Ayoub, S.; Majid, M.; Mir, M.S.; Soomro, A.B. Least Square-Support Vector Machine Based Brain Tumor Classification System with Multi Model Texture Features. Front Appl Math Stat 2023, 9, 1324054, doi:10.3389/FAMS.2023.1324054.

Analysis of Synthetic Data Utilization with Generative Adversarial Network in Flood Classification using K-Nearest Neighbor Algorithm

Wahyu Afriza, Mardhani Riasetiawan*, Dyah Aruming Tyas
Department of Computer Science and Electronics, Gadjah Mada University, Yogyakarta, Indonesia

Abstract**—Indonesia is a country with a tropical climate that has high rainfall rates and is supported by the uncertainty of weather and climate conditions. With the uncertainty of weather and climate as well as flood events, minimal predictive information on flooding, and the lack of availability of data on the causes of flooding, a comparison of synthetic data generation from the minimal data available from BMKG with synthetic data generation from Kaggle online platform data in the form of temperature and humidity data, rainfall, and wind speed from BMKG and annual rain data from Kaggle was analyzed. This research aims to obtain the results of data comparison analysis of synthetic data generation from different datasets with the benchmark of classification system results using K-Nearest Neighbor (KNN) and accuracy evaluation with Confusion Matrix. The research process uses climate data from the BMKG DI Yogyakarta Climatology Station within 20 months, the Geophysical Station within 12 months, and Kerala data with a range of 1901–2018. Synthetic data generation is done using the Conditional Tabular Generative Adversarial Network (CTGAN) model. CTGAN produces quite good data in terms of distribution and data differences if the original data is large and the synthetic data generated is small. The KNN classification system on the BMKG data experienced overfitting, as indicated by the accuracy value in the evaluation increasing in the range of 85–94% and the validation decreasing in the range of 89%–65%. This is because there is no uniqueness in the data and too little original data made into synthetics, which affects the difficulty of the classification system in identifying data that is quite different in distance and data values generated by CTGAN. In Kerala, the accuracy value on evaluation is in the range of 92–95%, and validation is in the range of 0.7–0.83%, with Classifier k1 being the most optimal system.**

Keywords—Classification; rainfall; synthetic data; KNN; GAN

I. Introduction

Indonesia is a tropical country located in the equatorial region with high rainfall. Today's climate change is affecting the weather and climate to the extent that high water discharge causes flooding. Water discharge can be caused by heavy rains with short or long durations, as in the rains that hit the Asian sub-continent with the deadliest floods that damage the environment, agricultural land, and basic health facilities [1]. A flood is a state of water inundation for a certain time, even though it is in an area that rarely floods with the support of rainfall and a long duration [2]. Floods can affect living things, wind pressure, temperature, watercolor, wind direction, humidity, and more, and have the potential to damage property and buildings [3], as well as adversely affect human health, the environment, cultural heritage, and economic activities [4].

The characteristics of heavy rainfall are strongly influenced by spatiotemporal patterns, space- and time-based models, and the amount of rainfall. Location: with damage intensity within the watershed, damage patterns (flooding from rivers, flooding from inland waters, sediment-related disasters, and other) vary depending on the distribution of rainfall. In order to implement effective flood control measures, it is important to understand the rainfall patterns that occur in the watershed and take countermeasures based on the characteristics of the associated hazards [5].

Some of the factors that affect the occurrence of floods are temperature, humidity, dew point temperature, wind speed, river flow volume, water level, and rainfall volume. The amount of rainfall is a major factor in the hydrological cycle process by monitoring the balance of freshwater and saltwater resources. Process of data acquisition can be done with the use of the Internet of Things based on data from the sensors used. Rainfall prediction or forecasting plays an important role in hydrological modeling and management of water resource issues such as flood warnings and real-time control of urban drainage systems [6].

In this research, a comparative analysis of the use of synthetic data in making a classification system based on the machine learning algorithm K-Nearest Neighbor (KNN) is carried out. Synthetic data generation is carried out due to the lack of availability and types of data features that can be obtained from BMKG Online Data (https://dataonline.BMKG.go.id/) and open data from online platforms for the classification of flood disaster events. The data used is BMKG data with rainfall data parameters, temperature and humidity, wind speed, and flood events as benchmarks for measuring and determining potential flood classes as well as monthly and annual flood data. This research is intended as an analysis of the use of synthetic data on climate data and natural disasters.

II. Research Methodology

A. System Needs Analysis

There are several stages in designing a classification system, including design, data preparation, training, and testing of a classification system based on the K-Nearest Neighbor (KNN) algorithm. The data used is BMKG data as training and validation data and Kerala data. Dataset creation includes downloading data and merging BMKG and BPBD data to become BMKG data with data that has a flood class label. The flood level data entry is in accordance with Table I. Then, the Kerala data underwent a download process without any additional processing.

TABLE I. FLOODING LEVEL IN YOGYAKARTA

No.	Flood Level	Flood High
1	Tidak/No	0 cm
2	Ringan/Low	< 100 cm
3	Tinggi/High	≥ 100 cm

BMKG training data has a time span of twenty months starting from January 2022 to August 2023; BMKG validation data has a time span of twelve months or one year starting from October 2022 to September 2023; and Kerala data in the form of rainfall data and annual flood classes has a time span of 1901–2018.

The BMKG training data will be divided into a 3:1 ratio for training and testing, so that 75% of the data will be used in the training process and 25% of the data will be used in the testing process, which can be used as confusion matrix-based evaluation results. The BMKG data has a total of 320 rainfall data points over a time span of twenty months. In the validation of BMKG data, the data is fully used as validation of the classification system results. Furthermore, Kerala data totals 118 data points, which will be divided into 100 training data points with the same division as BMKG data, namely 75% and 25%, and 18 data points as validation data from the Kerala classification system.

B. Synthetic Data Generation

Synthetic data is artificial data generated from the original data. Synthetic data can overcome the problems of data security, data confidentiality, unbalanced data, and others. The generative adversarial network works based on two neural networks: the discriminator and the generator [7]. GAN has several mathematical formulas for calculations. In GAN, there is a discriminator in Eq. (1), a generator in Eq. (2), and training for the discriminator and generator is shown in Eq. (3) and Eq. (4) [8].

$$L_D = Error(D(x),1) + Error\big(D\big(G(z)\big),0\big) \tag{1}$$

$$L_G = Error\big(D\big(G(z)\big),1\,\big) \tag{2}$$

$$V(G,D) = E_{x\sim P_{data}}\big[\log\big(D(x)\big)\big] + E_{z\sim P_z}[\log(1-D(G(z)))] \tag{3}$$

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \tag{4}$$

$$V(G,D^*) = E_{x\sim P_{data}}\big[\log\big(D^*(x)\big)\big] + E_{x\sim P_g}\big[\log\big(1-D^*(x)\big)\big] \tag{5}$$

C. Classification Method Implementation

After obtaining data from the source, the data will undergo a merging process between climate data and flood event data, then data cleaning will be carried out from unnecessary data or data with empty values and 8888 (unmeasured data) by manipulating the data using the median value.

The KNN method will use BMKG data which will be divided into train and test data. After defining the data, we will look for the minimum and maximum distance from the calculation of the train data distance and then the minimum distance from the calculation of the test data distance to the train data which will be assigned to several flood classes. The data will be classified into three flood classes, namely the No, Mild, and High classes as shown in Fig. 1.

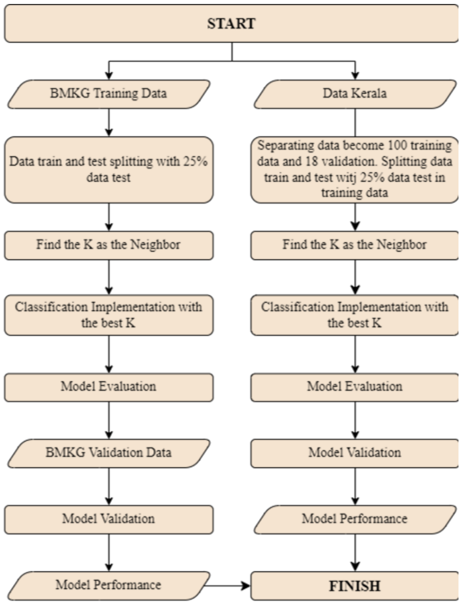

START
BMKG Training Data
Data Kerala
Data train and test splitting with 25% data test
Separating data become 100 training data and 18 validation. Splitting data train and test witj 25% data test in training data
Find the K as the Neighbor
Find the K as the Neighbor
Classification Implementation with the best K
Classification Implementation with the best K
Model Evaluation
Model Evaluation
BMKG Validation Data
Model Validation
Model Validation
Model Performance
Model Performance
FINISH

Fig. 1. Classification implementation.

D. Evaluation and Validation

The system evaluation process will be carried out by testing whether the system is able to classify rainfall, temperature, humidity, and wind speed data that can potentially flood into three flood classes. The measurement will be carried out by utilizing the Confusion Matrix Theory, which will compare the output of the system with the actual label of the data and will then produce accuracy, precision, Recall, and f1-score numbers according to Eq. (6) to Eq. (11) In the validation process, the same thing is done but with different data, namely data that has never experienced the train and test process.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$Precision = \frac{TP}{TP+TN} \tag{7}$$

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

$$FI - Score = \frac{1}{\frac{1}{recall}+\frac{1}{presisi}} \tag{9}$$

III. RESULT AND DISCUSSION

The datasets used in the research are from BMKG data and Kerala data, which are then divided into train, test, and validation data. BMKG data are gather from the filed sensor in the real situation. Kerala data will later be used in making a classification system with a validation process from different data and the training process that has been carried out. BMKG data that has been downloaded from the BMKG and BPBD

DIY online portals has a distribution of rain event data totaling 320 and 159 in BMKG training and validation data, respectively. The cleaning process is done by deleting the non-rain values in the BMKG data. In Kerala data, no cleaning was done because all data will be used.

After the non-rain data was eliminated, data manipulation was performed to fill in the values of the variable components in each column that were zero, null, and unmeasured except RR. The review for data manipulation was conducted only on the Tavg, RH_avg, and ff_avg features. Referring to Fig. 2, the data distribution on each feature except rainfall data (RR) is unevenly skewed with the category "skewed negative," or the mean value is lower than the median value of the data. In this situation, data manipulation using the median value aims to direct the distribution to a normal distribution. This was also done on the validation BMKG data.

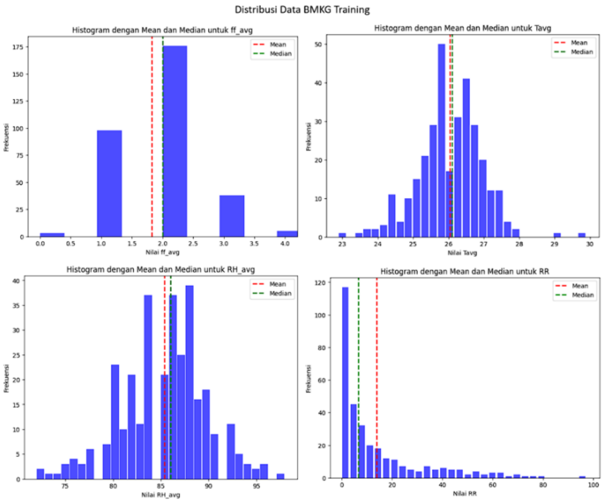


Fig. 2. Distribution of feature data.

After doing data manipulation on data rows that are 0, null, and unmeasured, as shown in Table II, In Kerala data, there are two flood classes, namely YES and NO as shown in Table III.

In the BMKG training data that has been processed, the data is not balanced between classes. If the process of making a classification system using this data is not followed, the result of the classification system will produce a poor system and undertraining, a situation when the classification system can recognize one class well but cannot recognize the other class, which tends to result in the classification system that has been made classifying data for the majority class [9]. This can be overcome by multiplying existing data with generative data or synthetic data.

The data generation process is carried out with CTGAN. This was done to prevent the potential for an undertrained classification system. The generative process was carried out with five experiments with different amounts of data. The flood class in the data will be converted into an integer or number, with 0 as no flood, 1 as a minor flood, and 2 as a high flood. The original flood data, with a total of six minor flood events and one high flood event, doubled without changes to two for the model calculation process, will be trained by GAN and generate synthetic data.

TABLE II. TOTAL BMKG DATA IN EACH CLASS

No.	Data	Flood	Case
1	BMKG Training	No	313
2		Mild	6
3		High	1
4	BMKG Validation	No	156
5		Mild	2
6		High	1

TABLE III. TOTAL KERALA DATA IN EACH CLASS

No.	Data	Flood	Case
1	Kerala Training	YES	52
2		NO	48
3	Kerala Validation	YES	8
4		NO	10

Furthermore, synthetic data was created for each data point. In BMKG data, synthetic data for mild flood and high flood classes is made into 30, 60, 90, 120, and 150 data points. In Kerala, synthetic data for yes and no classes was made into 150 of all the data and only 6 data samples. The distribution and differences between synthetic and real data can be seen in Fig. 3 to Fig. 14.

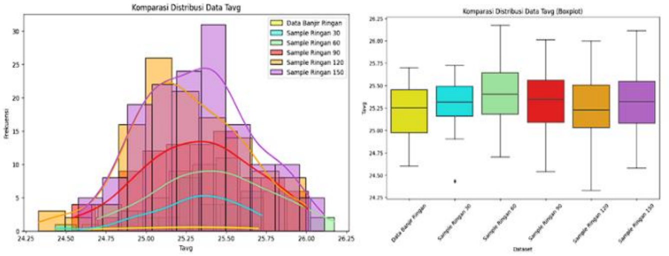


Fig. 3. Distribution of Tavg in mild class.

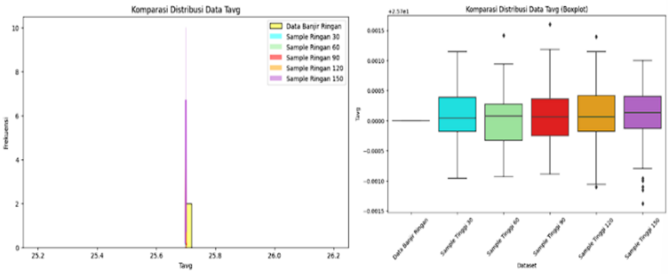


Fig. 4. Distribution of Tavg in high class.

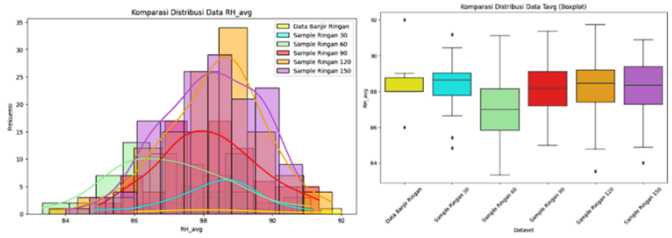


Fig. 5. Distribution of RH_avg in mild class.

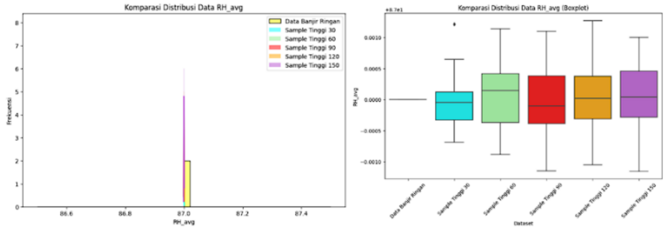


Fig. 6. Distribution of RH_avg in high class.

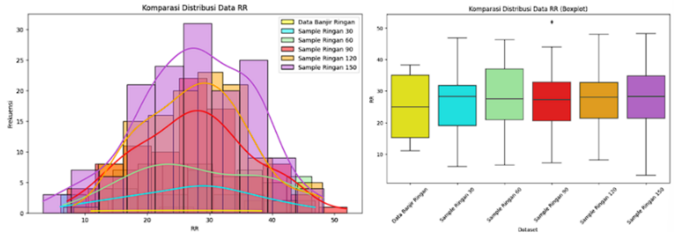


Fig. 7. Distribution of RR in mild class.

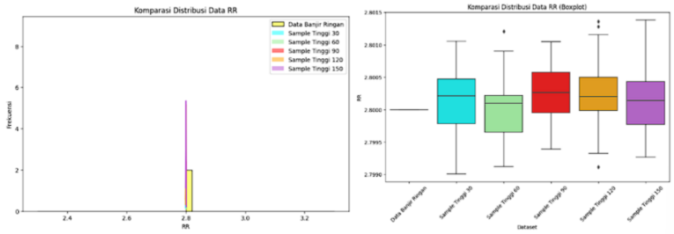


Fig. 8. Distribution of RR in high class.

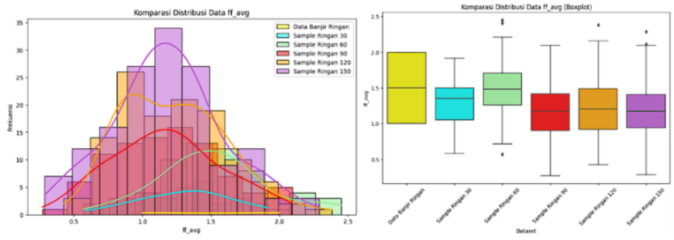


Fig. 9. Distribution of ff_avg in mild class.

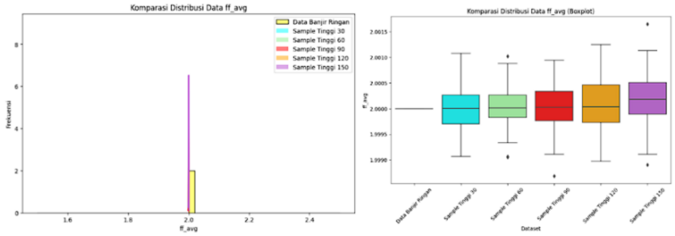


Fig. 10. Distribution of ff_avg in high class.

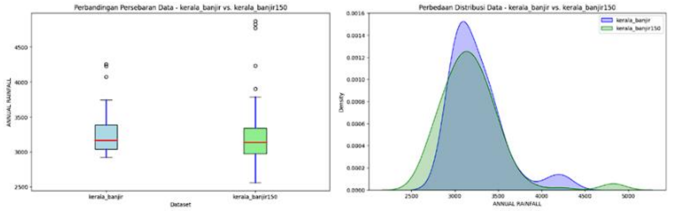


Fig. 11. Distribution of flood data in Kerala real data to 150

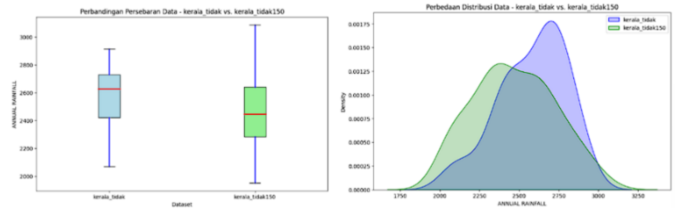


Fig. 12. Distribution of not flood data in Kerala real data to 150

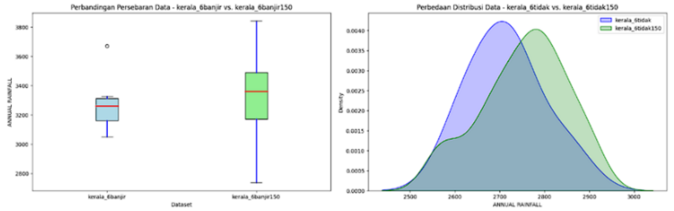


Fig. 13. Distribution of flood data in Kerala 6 real data to 150

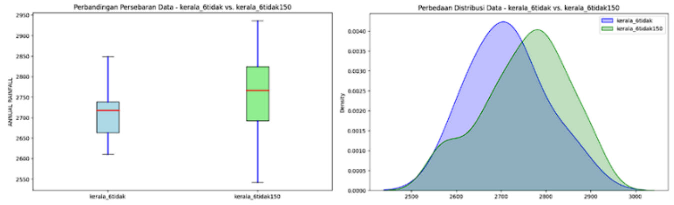


Fig. 14. Distribution of not flood data in Kerala 6 real data to 150

TABLE IV. TOTAL SYNTHETIC DATA OF BMKG DATA

No.	**Model**	**No Class**	**Mild Class**	**High Class**
1	*Classifier* 1	313	36	32
2	*Classifier* 2	313	66	62
3	*Classifier* 3	313	96	92
4	*Classifier* 4	313	126	122
5	*Classifier* 5	313	156	152

TABLE V. TOTAL SYNTHETIC DATA OF KERALA DATA

No.	Model	YES	NO
1	*Classifier k*	52	48
2	*Classifier k1*	202	198
3	*Classifier k2*	156	156

Before making a classification system in each experiment based on the data that shown in Table IV and Table V, the K value is determined as a consideration for determining neighbors in calcification using the Euclidean distance, which has a calculation formula as in Eq. (10) and Eq. (11).

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \quad (10)$$

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (11)$$

In the process of determining the value of K, the best value was found using repetition operations throughout the classifier experiment. Based on the accuracy results, a K value of 5 is obtained with the consideration of a good and consistent accuracy value, with an average accuracy value of 91.38% on BMKG data and 92.26% on Kerala data.

The classification system that has been created in each experiment will be evaluated to determine and assess the ability or performance of the system. In the system evaluation, each classifier is evaluated by utilizing the confusion matrix theory based on the formulas in Eq. (6) to Eq. (9). In the evaluation process, test data is used, which amounts to 25% of the total of all classes. The confusion matrix results of each classifier can be seen in Table VI for the BMKG data and Table VII for the Kerala data, as well as the comparison graph in Fig. 15.

TABLE VI. BMKG DATA CLASSIFICATION EVALUAION

No.	Model	Precission	Recall	F-1 score	Accuracy
1	*Classifier 1*	0.82	0.84	0.83	0.84
2	*Classifier 2*	0.91	0.91	0.91	0.91
3	*Classifier 3*	0.93	0.92	0.92	0.92
4	*Classifier 4*	0.93	0.92	0.93	0.92
5	*Classifier 5*	0.93	0.93	0.93	0.93

TABLE VII. KERALA DATA CLASSIFICATION EVALUATION

No.	Model	Precission	Recall	F-1 score	Accuracy
1	*Classifier k*	0.93	0.92	0.92	0.92
2	*Classifier k1*	0.91	0.90	0.90	0.90
3	*Classifier k2*	0.95	0.94	0.94	0.94

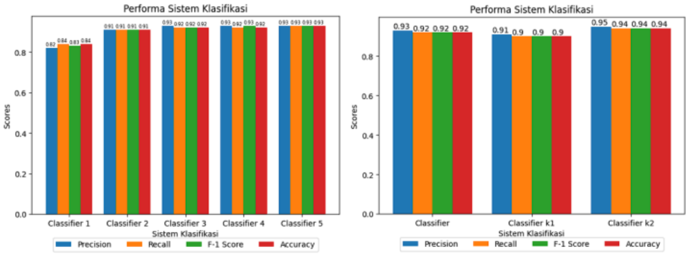


Fig. 15. Comparison of evaluation in each model.

Validation data on BMKG data has a total of 159 data points that have undergone data pre-processing. In the BMKG validation data, there are 156 non-flooding events, 2 minor floods, and 1 high flood. While in Kerala, the train data amounted to 18 data points, with 8 flood data points and 10 non-flood data points. In the validation test, the same process as the evaluation is carried out but with data that has never been trained and tested, utilizing the confusion matrix theory based on the formulas in Eq. (5), (6), (7), and (8) with the predicted data and the original label data. The results of the BMKG training data classification system validation can be seen in Table VIII, and the Kerala data system validation can be seen in Table IX. Then, the result comparison graph is in Fig. 16.

TABLE VIII. BMKG DATA CLASSIFICATION VALIDATION

No.	Model	Precission	Recall	F-1 score	Accuracy
1	*Classifier 1*	0.96	0.89	0.93	0.89
2	*Classifier 2*	0.96	0.86	0.91	0.86
3	*Classifier 3*	0.97	0.77	0.86	0.77
4	*Classifier 4*	0.97	0.74	0.83	0.74
5	*Classifier 5*	0.97	0.65	0.78	0.65

TABLE IX. KERALA DATA CLASSIFICATION VALIDATION

No.	Model	Precission	Recall	F-1 score	Accuracy
1	*Classifier k*	0.75	0.75	0.75	0.77
2	*Classifier k1*	0.78	0.88	0.82	0.83
3	*Classifier k2*	0.71	0.62	0.67	0.72

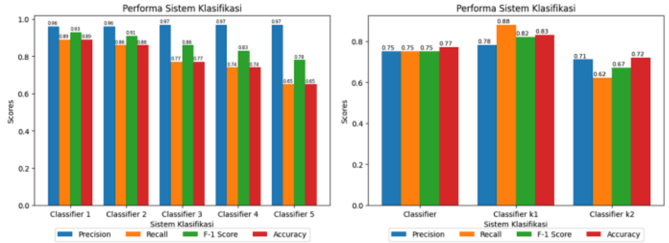


Fig. 16. Comparison of evaluation in each model

Referring to Fig. 17, the comparison of evaluation and validation results on BMKG data shows a significant difference in data, so the classification system from BMKG data has poor results, although Classifier 1 has the least difference. While in Kerala data, the evaluation and validation results are quite consistent in their improvement, with Classifier k1 being the most optimal, which is the creation of synthetic data from all the original Kerala data.

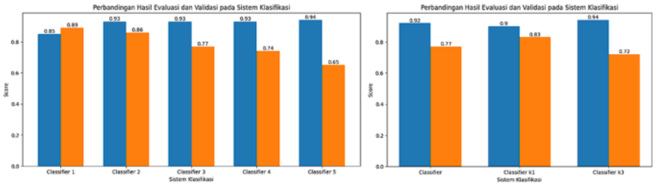


Fig. 17. Comparison of accuracy in evaluation and validation.

IV. CONCLUSION

The research indicates that the BMKG data, which includes temperature, humidity, rainfall, and wind speed features, does not have unique characteristics for each class. As a result, the classification system derived from this data suffers from overfitting, leading to imbalanced results between evaluation and validation. On the other hand, the Kerala data exhibits unique features for each class, which allows for a more accurate classification system. This system achieves an evaluation accuracy of 90-95% and a validation accuracy of 72-83%.

Synthetic data generated from Generative Adversarial Networks (GANs) can create a large amount of data from a small amount of original data. However, the quality of this synthetic data is dependent on the quantity of original and synthetic data used. This can affect the similarity of the synthetic data to the original data, and vice versa. In terms of the classifiers, Classifier 1 (with 30 sample data), Classifier 2 (with 60 data samples), and Classifier k1 show similar accuracy values on evaluation and validation. However, Classifier 3, Classifier 4, Classifier 5, and Classifier k3 exhibit a significant difference in accuracy values on evaluation and validation.

In summary, the BMKG data's lack of unique class characteristics leads to overfitting in the classification system, resulting in imbalanced evaluation and validation results. In contrast, the Kerala data, with its unique class characteristics, produces a more accurate classification system. Synthetic data, generated from GANs, can be highly useful, but the quality of this data depends on the quantity of original and synthetic data used. Finally, the classifiers show varying accuracy values on evaluation and validation, with Classifiers 3, 4, 5, and k3 exhibiting significantly different results compared to Classifiers 1, 2, and k1. The analysis of synthetic data utilization with Generative Adversarial Network (GAN) in flood classification using the K-Nearest Neighbor (KNN) algorithm is an innovative and promising research area. To further advance this work and contribute to the field, the following future work suggestions are proposed Explore and develop more advanced GAN architectures to generate synthetic flood-related data. Investigate different GAN variants, such as Wasserstein GANs or Progressive GANs, to improve the quality and diversity of synthetic data. Conduct a thorough investigation into the hyperparameters of both the GAN and KNN algorithms to optimize their performance. This includes tuning learning rates, batch sizes, and other relevant parameters to achieve better results in terms of classification accuracy and computational efficiency. Extend the analysis by incorporating and comparing the performance of other machine learning models for flood classification. This could include algorithms like Support Vector Machines (SVM), Decision Trees, or Random Forests. A comprehensive comparison will provide insights into the strengths and weaknesses of different approaches. Test the proposed framework on a broader range of datasets to evaluate its generalizability. This includes datasets from different geographical locations, varied environmental conditions, and various types of flooding scenarios. Assess the robustness of the model across different contexts. By addressing these future work areas, the research can make significant contributions to the field of flood classification, synthetic data generation, and the intersection of GANs and KNN algorithms.

REFERENCES

[1] Babar, M., Rani, M. and Ali, I., 2022, November. A Deep learning-based rainfall prediction for flood management. In 2022 17th International Conference on Emerging Technologies (ICET) (pp. 196-199). IEEE.

[2] Al Kindhi, B., Triana, M.I., Yuhana, U.L., Damarnegara, S., Istiqomah, F. and Imaaduddiin, M.H., 2022, November. Flood Identification with Fuzzy Logic Based on Rainfall and Weather for Smart City Implementation. In 2022 IEEE International Conference on Communication, Networks, and Satellite (COMNETSAT) (pp. 67-72). IEEE.

[3] Khan, T.A., Shahid, Z., Alam, M., Su'ud, M.M. and Kadir, K., 2019, December. Early flood risk assessment using machine learning: A comparative study of svm, q-svm, k-nn, and lda. In 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS) (pp. 1-7). IEEE.

[4] Panganiban, E.B. and Cruz, J.C.D., 2017, November. Rainwater level information with flood warning system using flat clustering predictive technique. In TENCON 2017-2017 IEEE Region 10 Conference (pp. 727-732). IEEE.

[5] Hoshino, T. and Yamada, T.J., 2023. Spatiotemporal classification of heavy rainfall patterns to characterize hydrographs in a high-resolution ensemble climate dataset. Journal of Hydrology, 617, p.128910.

[6] Adaryani, F.R., Mousavi, S.J. and Jafari, F., 2022. Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM, and CNN. Journal of Hydrology, 614, p.128463.

[7] Habibi, O., Chemmakha, M., & Lazaar, M. (2023). Imbalanced tabular data modelization using CTGAN and machine learning to improve IoT Botnet attacks detection. Engineering Applications of Artificial Intelligence, 118, 105669.

[8] Kiran, A. and Kumar, S.S., 2023, March. A Comparative Analysis of GAN and VAE based Synthetic Data Generators for High Dimensional, Imbalanced Tabular data. In 2023 2nd International Conference for Innovation in Technology (INOCON) (pp. 1-6). IEEE.

[9] Karimi, Z., 2021. Confusion Matrix. Encycl. Mach. Learn. Data Min., no. October, pp.260-260.

[10] Freiesleben, T. and Grote, T., 2023. Beyond generalization: a theory of robustness in machine learning. Synthese, 202(4), p.109. Karimi, Z., 2021. Confusion Matrix. Encycl. Mach. Learn. Data Min., no. October, pp.260-260.

[11] Freiesleben, T. and Grote, T., 2023. Beyond generalization: a theory of robustness in machine learning. Synthese, 202(4), p.109.

A Piano Single Tone Recognition and Classification Method Based on CNN Model

Miaoping Geng, Ruidi He*, Ziyihe Zhou
The Conservatory of Music, Hebei Institute of Communications, Shijiazhuang, Hebei 050010, China

***Abstract*—In order to improve the recognition and classification effect of piano single tone, this paper combines the CNN (Convolutional Neural Networks) model to construct the piano single tone recognition and classification model, and equalizes the uniformly irradiated parabolic tone transmission hardware. In this paper, the analytic method is used to calculate the direction diagram of the tone transmission hardware, and the analytical expression for calculating the gain of the tone transmission hardware is obtained. Moreover, this paper gives the calculation and analytical expression of the hardware gain of the tone transmission in the main lobe, and obtains the calculation method of the relative position of the two tone transmission hardware by using the conversion relationship between the global coordinate and the local coordinate. Finally, the variation law of the received power with the azimuth/elevation angle of the receiving tone transmission hardware and the incident high-power microwave frequency is given. The experimental study shows that the piano single tone recognition and classification method based on CNN model proposed in this paper can play an important role in piano single tone recognition. This article improves the note recognition algorithm for piano music by combining note features with frequency spectrum to obtain note spectrum, which improves the accuracy of audio classification recognition.**

Keywords—CNN model; piano; single tone recognition; classification

I. Introduction

For the interpretation of a beautiful piano piece, if the performer is the driver of the soul of the piece, then a piano with a beautiful tone is the carrier of the soul. Therefore, it is particularly important to understand and study the quality of the piano. The quality of a piano can be judged from six aspects: tone, touch, tuning stability, durability, appearance and tension. However, for players, tone and touch are undoubtedly the most important. It is not an easy task to make a correct judgment on the sound quality of a piano. First of all, it needs to ensure the piano intonation. In the case of intonation, the pronunciation of the bass region should be strong and powerful, the sound should be extended enough, and the pronunciation should not be short or weak. The mid-range requires that the extension of the sound should be as long as possible, the timbre should be beautiful and soft, and the pronunciation should not be dull or blunt. The pronunciation of the high-pitched area should be bright and clear, not too gorgeous or impure, and no reverberation is required. When the timbres of the three sound zones are satisfactory, it should be noted that the transition of the three sound zones should be natural, the timbre should be unified, and the volume ratio should be coordinated. The so-called tactile sensation refers to the responsive ability of the keyboard and the action mechanism to transmit the player's playing force to the strings. The touch of the piano should be comfortable for beginners and comfortable for accomplished players.

Music information signals themselves belong to fuzzy signals and require strict mathematical models to describe them. Traditional information processing methods are difficult to solve the fuzzy situation of music signals. Therefore, some studies use intelligent information analysis and processing methods such as fuzzy systems, neural networks, expert systems, and genetic algorithms to process music information [1].

Fuzzy system has the advantages of not needing precise mathematical model, easy to use human experience knowledge, nonlinear, robust and so on. Since there are many ambiguities in music information, language and thinking, the use of fuzzy sets to describe the characteristics of music, and the use of fuzzy logic and fuzzy reasoning for feature analysis and identification, should be said to be the closest to the cognitive process of people's music. At present, many studies have proved that the fuzzy system is an effective method for the study of music information. In the automatic chord analysis system of literature [2], the process of listening, feeling and understanding music all use the membership function of the fuzzy set, and the basic membership function The function is designed according to music theory and can be modified by a few simple parameters, which facilitates the study of music information from both music theory and human perception.

In order to improve the effect of piano single tone recognition and classification, this paper combines the CNN model to build a piano single tone recognition and classification model to improve the efficiency of intelligent processing of piano tones.

The contribution of this article is as follows: Section I is the introduction. Related work is given in Section II. Calculation and Analysis of front door of coupling quantity of piano tone transmission is given in Section III. Section IV delves in to the empirical model, results, analysis and discussion. Finally, Section V concludes the paper. This article improves the note recognition algorithm for piano music by combining note features with frequency spectrum to obtain note spectrum, which improves the accuracy of audio classification recognition.

*Corresponding Author.

II. Related Work

Neural network has the advantages of distributed information storage, parallel processing, self-organization, and self-learning, the existing research results use two types of neural networks [3]: (1) Multilayer perceptron (BP network): Using a layered neural network can make the system not only start from the notes, but also from the overall structure. Highly comprehensive processing of music information. The hierarchical recurrent neural network of [4], the first layer is for each measure, and the second layer is for each note. For a rhythm-regular piece like a waltz, both melody and rhythm characteristics can be taken into account. In addition, the use of Makov chains can also make the system predictive. (2) Hopfield network (feedback network): Research in [5] uses the associative memory function of the Hopfield network to recombine the existing melody data to achieve the purpose of composition. The expert system is suitable for embedding a large amount of music knowledge into the computer system, and at the same time enables the system to flexibly use this knowledge to make judgments, and has the ability to acquire and increase knowledge, and is best at simulating human experts to solve complex problems. Study [6] used 26 rules to complete the automatic playing expert system according to the performance characteristics of performers. Study [7] realized an expert system that automatically discriminates Bach's musical style. In the processing of musical information, intelligent methods play an important role. The choice of these methods should mainly be based on specific needs. For example, the analysis of the characteristics of music harmony, timbre, etc. is often inseparable from a fuzzy system that has a strong ability to describe music information. Automatic playing systems usually choose expert systems that are easy to acquire knowledge. Some auxiliary composing systems use expert systems, while others use expert systems. The neural network is used to simulate human image thinking, so as to compose music according to certain requirements [8].

Study [9] argues: "The typical sound quality of an instrument should be attributed to the relative intensities of all harmonics in a definite or relatively definite region in the musical scale." Although this view is not entirely correct, it points out that the sound quality of an instrument is the importance of harmonic amplitudes to the sound quality of musical instruments. In terms of using computer to synthesize piano sound and improving the sound quality of piano sound through computer processing simulation, literature in [10] pointed out that harmonics are an important factor of sound quality, but this paper discusses how to improve the sound quality from the perspective of harmonic amplitude and phase changing with time, piano sound. Study in [11] believes that the harmonic amplitude is an important factor that constitutes the sound quality of the piano, and then uses the method of simulating multiple strings to reasonably adjust the frequency spectrum of the piano to study how to improve the sound of the piano.

When the hammers hit the strings, the piano sound reaches a peak of vibration soon after a brief onset. From a musical point of view, it is better to have a shorter time in this stage. If the time is too long, the sound will appear soft and lack the feeling of rigidity; but it should not be too short, otherwise the sound will have a stiff feeling. When the peak of the sound amplitude is reached, there will be a rapid decay. This period can be divided into two stages: early decay 2 and late decay 3. Of course, the decay speed of this stage will vary greatly with different keys and different percussion strengths [12]. An excellent pianist can well control the percussion intensity and time during the performance, so that the piano can make a full sound and get a good timbre effect. The strings are restrained by the sound felt and the vibration is rapidly attenuated. For the treble keys of a piano, these processes are less complicated, and the time domain graph of the treble keys looks like a straight line with only a sloping downward trend. The time domain characteristics of all these piano sounds are closely related to the piano hardware itself, such as felt, hammer shape, soundboard, etc. That is to say, the hardware of the piano itself is a very important part of the time domain characteristics of the piano sound. determinants [13].

Regarding the influence of the harmonic amplitude of the piano on the sound quality, the literature [14] first recorded the piano, and then took one of the signals, and performed simple processing on the harmonic amplitude, for example, the harmonic amplitude was formed proportionally, or the harmonic amplitude was decreased according to the law form a set of synth sounds, etc. Then compare these synthetic sounds with the original recording of the piano, and finally find that the sound quality after such simple processing is not as good as the original sound, at most only close to the original recording. , processing the harmonic amplitude of the piano is not a simple process, and it also shows that the composition of the harmonic amplitude does affect the sound quality of the piano sound to a great extent. In addition to studying the influence of harmonic amplitude on sound quality, study in [15] found the law of piano harmonics changing with time, that is, the law that different harmonics have different time delays. For example, the first harmonic and the second harmonic appear at different times, and there is a small delay, which means that the second harmonic will come later than the first harmonic. And as the harmonic order increases, their delay time increases.

Using a computer to analyze piano music signals, the processed music information must first be digitally processed, that is, convert the analog music signals collected by the recording equipment into digital music signals. In this process, two aspects are mainly considered: conversion accuracy and operation efficiency. The A/D conversion accuracy is realized by the number of bits of the A/D conversion device; the operation efficiency is related to the digitization accuracy and sampling rate of the music. However, if the data accuracy and sampling rate are too low, it will cause relatively large waveform distortion in the digitization process of the signal [16].

The collected piano music will be studied from the perspectives of physical acoustics, rhythm, fast Fourier transform, wavelet analysis, etc., to study the specific extraction methods and analysis methods of different musical features, and complete the design of pitch, duration, intensity and other feature extraction methods. and its computer implementation. The extraction of musical features provides a practical basis for the design of subsequent piano evaluation systems [17].

The keystroke sensitivity of the piano keys, that is, the response frequency of the keyboard, is one of the main parameters affecting the touch feeling of the piano keys, and it is also a problem that piano manufacturers are concerned about at present. By analyzing whether the response times of the piano keys can reach the standard, the quality of the percussion performance of the piano keys can be judged. To test this parameter, we must first ensure that the strength of hitting the piano keyboard is the same, the interval is even, and the frequency is adjustable, so that accurate test results can be obtained. In terms of testing algorithms, the keystroke sensitivity of piano keys can be effectively detected through waveform normalization, endpoint detection, and single-note separation [18].

III. Calculation and Analysis of Front Door Coupling Quantity of Piano Tone Transmission Hardware Port

On the basis of studying the characteristics of high-power microwaves, in order to quantitatively obtain the coupling number of high-power microwaves through the front door to the port of the piano tone transmission hardware, it is necessary to study the calculation method of the coupling amount of the front door.

A. Description of Calculation Method of Front Door Coupling

The HPM (High Performance Computing) generated in the high-power microwave transmitter transmits the hardware radiation through piano tone. As shown in Fig. 1, piano tone transmission hardware 1 is used as receiving, and piano tone transmission hardware 2 is used as piano tone transmission hardware for transmitting HPM.

When the distance between the received piano tone transmission hardware and the HPMpiano tone transmission hardware is R, the power density S from the HPM to the receiving piano tone transmission hardware is:

$$S = \frac{P_t G_t\left(\theta_2, \varphi_2\right)}{4\pi R^2}\left[\mathrm{W} / \mathrm{m}^2\right] \tag{1}$$

The content calculated by Formula (1) is the power density of the far-field field, and the condition for satisfying the far-field field is:

$$R \geq 2d^2 / \lambda \tag{2}$$

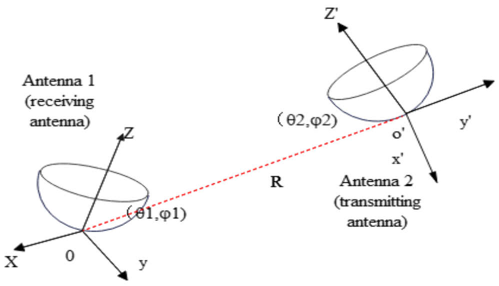


Fig. 1. The relative position of the two piano tone transmission hardware.

According to the Frith transmission formula, the power P_r received as the received piano tone transmission hardware is expressed as:

$$P_r = A_e\left(\theta_1, \varphi_1\right) \cdot S \tag{3}$$

The power received by the piano tone transmission hardware can be obtained from Formulas (1) and (3). In Formulas (1) and (3), the specific gain and effective receiving area of the two piano tone transmissions hardware need to be obtained.

B. Achieve Method of the Gain of the Piano Tone Transmission Hardware

For any form of piano tone transmission hardware, some commercial electromagnetic simulation software (HFSS (High Frequency Structure Simulator), FEKO etc.) can be used to establish the model of piano tone transmission hardware. The direction diagram and gain of piano tone transmission hardware are obtained through simulation.

The parabolic piano tone transmission hardware belongs to a class of piano tone transmission hardware with symmetrical structure, and its direction map and gain can be obtained by analytical method.

As shown in Fig. 2, for the larger rotating paraboloid piano tone transmission hardware with uniform illumination aperture shown in Fig. 2(a), the radiated electromagnetic waves. In essence, it can be equivalent to the electromagnetic wave radiated by the circular aperture (same diameter as the paraboloid) on a metal plate with an infinite size irradiated by a uniform plane wave as shown in Fig. 2(b).

For the equivalent model, the method for obtaining the far-field pattern can be based on the Huygens principle. According to this method, the normalized field strength pattern $E(\theta)$ obtained is shown in Formula (4).

$$E(\theta) = \frac{J_1\left[\pi d_\lambda \sin\theta\right]}{\sin\theta} \cdot \frac{2}{\pi d_\lambda} \tag{4}$$

In the formula, J_1 represents the first-order Bessel function, and θ refers to the angle relative to the focal axis.

d_λ is the wavelength number of the diameter d of the circular mouth, which can be expressed as

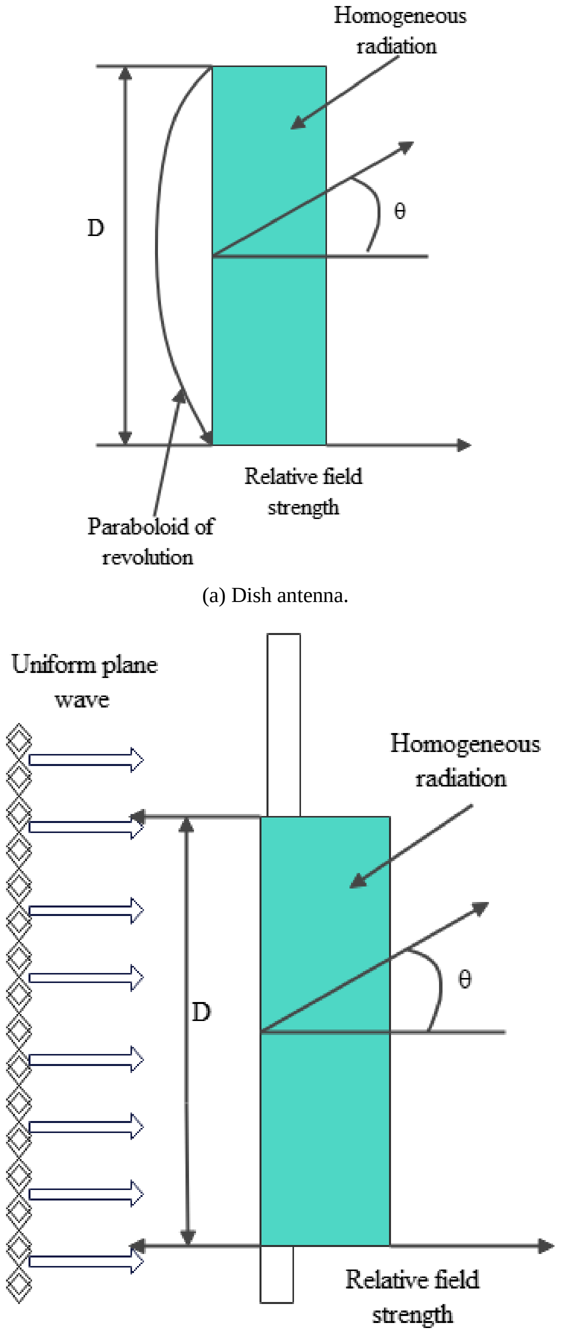
$$d_\lambda = d / \lambda \tag{5}$$

The n -order Bessel function can be expressed as:

$$J_n(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(n\theta - z\sin\theta)d\theta \tag{6}$$

The expression of the first-order Bessel function is shown in Formula (7).

$$J_1(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(\theta - z\sin\theta)d\theta \tag{7}$$



(a) Dish antenna.

(b) Equivalent model.

Fig. 2. The paraboloid model of revolution and its equivalent model.

Fig. 3 is the normalized field strength pattern when the center frequency of the parabolic piano tone transmission hardware is $f = 10\text{GHz}$ and the aperture is $d = 1\text{m}$.

Formula (8) represents the formula for obtaining the directivity coefficient of the parabolic piano tone transmission hardware:

$$D(\theta,\varphi) = (\tilde{E}(\theta,\varphi))^2 \cdot \frac{4\pi}{\theta_{HP1}\theta_{HP2}} \quad (8)$$

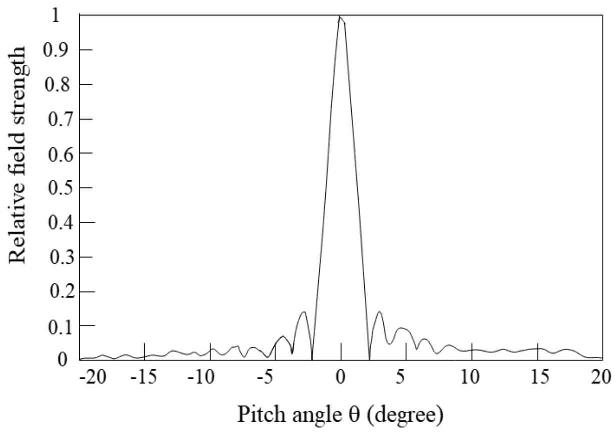


Fig. 3. Normalized direction diagram.

In the formula, $\tilde{E}(\theta,\varphi)$ is the normalized electric field intensity in the direction away from the piano tone transmission hardware, and θ_{HP1} and θ_{HP2} are the half-power beam widths of the main lobes of the piano tone transmission hardware in the two main planes.

The gain of the piano tone transmission hardware in a certain direction θ and φ is defined as:

$$G(\theta,\varphi) = \eta D(\theta,\varphi) \quad (9)$$

In the formula, η is the efficiency factor of piano tone transmission hardware.

When one piano tone transmission hardware (transmit/receive) is located in the main lobe of another piano tone transmission hardware (receive/transmit), the gain of piano tone transmission hardware can be calculated by the following method.

When the receiving piano tone transmission hardware is in the main lobe of the HPMpiano tone transmission hardware, the gain of the transmitting piano tone transmission hardware can be expressed as:

$$G_t = \varepsilon_a \frac{4\pi}{\lambda^2} A_{pt} \quad (10)$$

In the formula, ε_a is the aperture efficiency of the piano tone transmission hardware, and A_{pt} is the actual aperture area of the piano tone transmission hardware.

Similarly, when the HPM transmitting piano tone transmission hardware is located in the main lobe of the receiving piano tone transmission hardware, the gain of the receiving piano tone transmission hardware can be expressed as:

$$G_r = \varepsilon_a \frac{4\pi}{\lambda^2} A_{pr} \quad (11)$$

In the formula, A_{pr} is the actual aperture area of the receiving piano tone transmission hardware.

C. Calculation Method of Effective Receiving Area of Piano Tone Transmission Hardware

The effective receiving area $A_e(\theta,\varphi)$ of the piano tone transmission hardware is related to the gain of the receiving piano tone transmission hardware and the polarization mismatch coefficient, which can be expressed as:

$$A_e(\theta,\varphi)=\rho\cdot G_r(\theta,\varphi)\cdot A_d \tag{12}$$

In the formula, ρ is the polarization mismatch coefficient, the value is between $0\sim 1$, and A_d is the integral effective area of piano tone transmission hardware. When calculating the integral effective area of the piano tone transmission hardware, it is necessary to consider the relationship between the frequency f_t of the HPM transmitting piano tone transmission hardware and the center frequency f_r of the receiving piano tone transmission hardware. When $f_t < f_r$, it is called down-band coupling. When $f_t \approx f_r$, it is called in-band coupling. When $f_t > f_r$, it is called on-band coupling. The up-band coupling and the down-band coupling are collectively referred to as out-of-band coupling.

When calculating the in-band coupling, λ is not only the working wavelength, but also the wavelength corresponding to the center frequency of the receiving piano tone transmission hardware. When discussing the case of out-of-band coupling, the wavelength λ_r corresponding to the center frequency of the receiving piano tone transmission hardware should be used.

In the calculation model of the front door coupling amount, it is necessary to calculate the gain $G_t\left(\theta_2,\varphi_2\right)$、$G_r\left(\theta_1,\varphi_1\right)$ of the transmitting piano tone transmission hardware and the receiving piano tone transmission hardware. Because when calculating the direction map and gain of the piano tone transmission hardware, the angle is the relative coordinate system of the piano tone transmission hardware itself. Therefore, when calculating $\left(\theta_1,\varphi_1\right)$, the coordinates of the transmitting piano tone transmission hardware should be converted into the relative coordinate system of the receiving piano tone transmission hardware. When calculating $\left(\theta_2,\varphi_2\right)$, the coordinates of the receiving piano tone transmission hardware should be converted into the relative coordinate system of the transmitting piano tone transmission hardware.

For any point Q on the piano tone transmission hardware, its coordinate in the local coordinate system of the piano tone transmission hardware is $\left(x',y',z'\right)$, and its coordinate in the overall coordinate system is (x, y, z). Then, the overall coordinates (x, y, z) of the point Q can be regarded as obtained by the local coordinate $\left(x',y',z'\right)$ through two rotation transformations and one translation transformation.

1) When the x angle is rotated counterclockwise around the θ -axis, the corresponding coordinate transformation matrix is:

$$T_1=\begin{bmatrix}\cos\theta & 0 & \sin\theta & 0\\ 0 & 1 & 0 & 0\\ -\sin\theta & 0 & \cos\theta & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{13}$$

2) When the angle φ is rotated counterclockwise around the z-axis, the corresponding coordinate transformation matrix is:

$$T_2=\begin{bmatrix}\cos\varphi & -\sin\varphi & 0 & 0\\ \sin\varphi & \cos\varphi & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{14}$$

3) When $\left(x_0,y_0,z_0\right)$ is translated along the x-axis, y-axis, and z-axis, the corresponding coordinate transformation matrix is:

$$T_3=\begin{bmatrix}1 & 0 & 0 & x_0\\ 0 & 1 & 0 & y_0\\ 0 & 0 & 1 & z_0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{15}$$

4) In the global coordinates, the corresponding relationship between the point Q (x, y, z) and the point in the local coordinates $Q'\left(x',y',z'\right)$ is:

$$\begin{bmatrix}x\\ y\\ z\\ 1\end{bmatrix}=T\begin{bmatrix}x'\\ y'\\ z'\\ 1\end{bmatrix} \tag{16}$$

The transformation matrix T is:

$$\begin{aligned}T&=T_3\cdot T_2\cdot T_1\\ &=\begin{bmatrix}1 & 0 & 0 & x_0\\ 0 & 1 & 0 & y_0\\ 0 & 0 & 1 & z_0\\ 0 & 0 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}\cos\varphi & -\sin\varphi & 0 & 0\\ \sin\varphi & \cos\varphi & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}\cos\theta & 0 & \sin\theta & 0\\ 0 & 1 & 0 & 0\\ -\sin\theta & 0 & \cos\theta & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\\ &=\begin{bmatrix}\cos\theta\cos\varphi & -\sin\varphi & \sin\theta\cos\varphi & x_0\\ \cos\theta\sin\varphi & \cos\varphi & \sin\theta\sin\varphi & y_0\\ -\sin\theta & 0 & \cos\theta & z_0\\ 0 & 0 & 0 & 1\end{bmatrix}\end{aligned} \tag{17}$$

In the actual simulation calculation, the overall coordinates of a certain point on the piano tone transmission hardware are usually known. Therefore, it is necessary to convert the point coordinates in the overall coordinates into the local coordinate system, that is,

$$\begin{bmatrix}x'\\ y'\\ z'\\ 1\end{bmatrix}=T^{-1}\begin{bmatrix}x\\ y\\ z\\ 1\end{bmatrix} \tag{18}$$

Among them, the transformation coordinate T^{-1} is:

$$T^{-1} = \left(T_3T_2T_1\right)^{-1} = T_1^{-1}T_2^{-1}T_3^{-1} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi & 0 & 0 \\ -\sin\varphi & \cos\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -x_0 \\ 0 & 1 & 0 & -y_0 \\ 0 & 0 & 1 & -z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \cos\theta\cos\varphi & \cos\theta\sin\varphi & -\sin\theta & -x_0\cos\theta\cos\varphi - y_0\cos\theta\sin\varphi + z_0\sin\theta \\ -\sin\varphi & \cos\varphi & 0 & x_0\sin\varphi - y_0\cos\varphi \\ \sin\theta\cos\varphi & \sin\theta\sin\varphi & \cos\theta & -x_0\sin\theta\cos\varphi - y_0\sin\theta\sin\varphi - z_0\cos\theta \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{19}$$

D. Conversion Relationship between Received Power and Field Strength

Since the calculation model of the coupling quantity of the front door obtains the received power, the required parameter is the field strength value when designing the waveguide plasma limiter. Therefore, it is necessary to obtain the conversion relationship between the power and the field strength.

The conversion relationship between field strength and power is shown in Formula (20).

$$E_{[\mathrm{dB}(\mu V/m)]} = P_{[\mathrm{dBm}]} + AF + 107 \tag{20}$$

In the formula, AF is the piano tone transmission hardware factor.

In Formula (20), the received power can be obtained from the calculation model of the front door coupling, and the unknown parameter is the piano tone transmission hardware factor.

The piano tone transmission hardware factor is defined as the ratio between the electric field E and the piano tone transmission hardware terminal voltage V_L, namely:

$$AF = \frac{E}{V_L} \tag{21}$$

It is converted to decibels and expressed as,

$$E_{[\mathrm{dB}(\mu V/m)]} = V_{[\mathrm{dB}(\mu V)]} + AF_{\left[\mathrm{dB}\left(m^{-1}\right)\right]} \tag{22}$$

The effective receiving area A_e of the piano tone transmission hardware can be described as the ratio of the output power P_{out} of the piano tone transmission hardware to the incident power density P_d of the electromagnetic wave, namely:

$$A_e = \frac{P_{\mathrm{out}}}{P_d} \tag{23}$$

The effective area of the piano tone transmission hardware can also be calculated according to Formula (24).

$$A_e = \frac{G_r\lambda^2}{4\pi} \tag{24}$$

The output power of the piano tone transmission hardware can be expressed as:

$$P_{out} = \frac{V_L^2}{Z} \tag{25}$$

The power density of the incident electromagnetic wave is:

$$P_d = \frac{E^2}{120\pi} \tag{26}$$

From Formula (23) (26), we can get:

$$\frac{V_L^2}{Z} = \frac{E^2}{120\pi} \times \frac{G_r\lambda^2}{4\pi} \tag{27}$$

Therefore, the piano tone transmission hardware factor can be expressed as:

$$AF = \frac{E}{V_L} = \sqrt{\frac{480\pi^2}{Z\lambda^2 G_r}} \tag{28}$$

For the 50 ohm piano tone transmission hardware system, the piano tone transmission hardware factor can be expressed as:

$$AF = \frac{9.37}{\lambda\sqrt{G_r}} \tag{29}$$

It is expressed in decibels as:

$$AF = 19.8 - 20\times\log(\lambda) - 10\times\log\left(\mathrm{G_r}\right) \tag{30}$$

E. Scope of Application of the Calculation Model of Front Door Coupling

The distance from the piano tone transmission hardware is different, so that the field around the piano tone transmission hardware can be divided into three areas: the induction field area, the radiation near field area (Fresnel area) and the radiation far field area (see Fig. 4).

In the induction field area, piano tone transmission hardware does not radiate power. In the radiation near-field area, where it is close to the piano tone transmission hardware, it is difficult to distinguish the main lobe and side lobe of the pattern. As the distance becomes farther, the envelope of the pattern becomes clearer.

The direction map of the piano tone transmission hardware in the far field almost no longer changes with the distance. The measurement results at infinity from the piano tone transmission hardware are compared with the measurement results at a certain distance from the piano tone transmission hardware. If the difference between the two results is acceptable in engineering, the distance is called the Rayleigh distance, and here it is called the inner boundary of the far field.

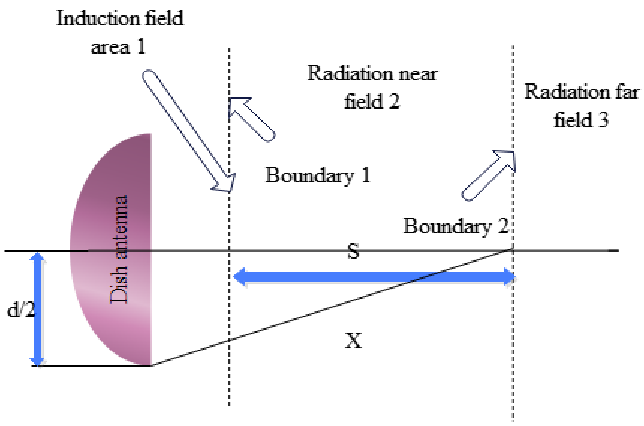


Fig. 4. Division of the piano tone transmission hardware field.

It can be seen from the measurement results that at the position where the Rayleigh distance is located, the distance from the center of the aperture to the observation point is:

$$x - s = \frac{\lambda}{16} \tag{31}$$

From the Pythagorean Theorem, we can get:

$$\left(\frac{d}{2}\right)^2 + s^2 = x^2 \tag{32}$$

Combining Formulas (31) and (32), the expression for the Rayleigh distance S is obtained as:

$$s = \frac{2d^2}{\lambda} - \frac{\lambda^3}{2048} \approx \frac{2d^2}{\lambda} \tag{33}$$

Therefore, the analytical calculation method of the parabolic piano tone transmission hardware pattern obtained in this paper is suitable for the far field.

IV. The Piano Single Tone Recognition and Classification Method based on CNN Model

A. Empirical Model

The system environment adopts Windows 10+Python 3.5, and the model training adopts a more concise and effective TFLearn model library. Due to limitations in conditions, the TensorFlow used in this article is the CPU (Central Processing Unit) version (the training effect may be better using the GPU (graphics processing unit) version), and the training status information comes from the auxiliary function output of TFLearn.

In order to realize the effective identification and classification of piano single tones, this paper combines the CNN algorithm to carry out research, and adopts the note-level method based on convolutional neural network (CNN) for multi-tone steel piano tone frequency recognition. The model flow is shown in Fig. 5.

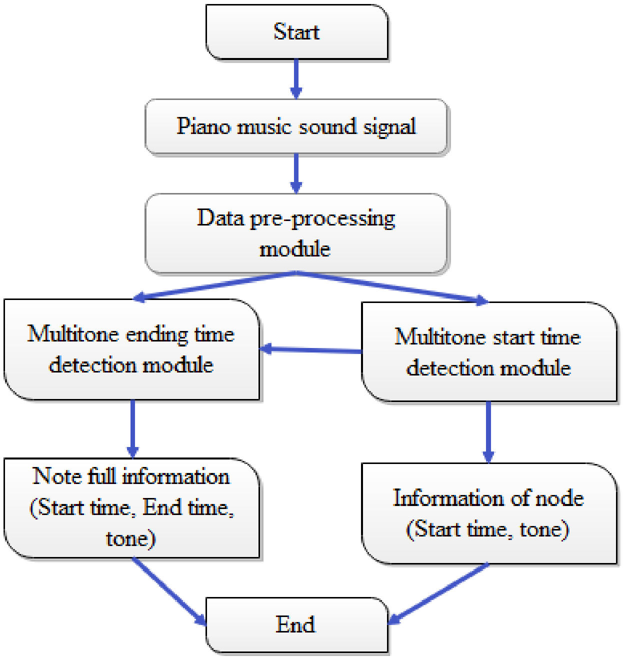

Start
Piano music sound signal
Data pre-processing module
Multitone ending time detection module
Multitone start time detection module
Note full information (Start time, End time, tone)
Information of node (Start time, tone)
End

Fig. 5. Structure design of the algorithm system.

Compare the model proposed in this article with the method proposed in study [4], and explore their performance in piano single tone recognition and classification. Quantitative evaluation was conducted using expert evaluation methods to obtain the results shown in Table I.

B. Results

The multi-tone onset time model detects the onset time points of multiple notes, and the note onset time points can see obvious frequency variation characteristics on the spectrogram, as shown in Fig. 6. The CQT video feature map of the F#4 notes has obvious edge mutation, which is very suitable for CNN.

Using FEKO software, the parabolic piano tone transmission hardware is modeled. The piano tone transmission hardware size is: the aperture is 1m, and the center frequency is 5GHz. The normalized direction diagram of the piano tone transmission hardware is obtained as shown in Fig. 7.

The hardware gain analysis calculation method of piano tone transmission in the main lobe is compared with the results of FEKO software simulation. The FEKO software is applied to model the piano tone transmission hardware of the speaker, and calculate the stereo pattern and gain of the piano tone transmission hardware when the frequency is 10GHz, as shown in Fig. 8.

After the above model is constructed, the effect of the model is verified by the piano single tone recognition and classification method based on the CNN model, and the verification is carried out through multiple sets of piano single tone recognition experiments, and the classification results are counted, and the schematic results shown in Fig. 9 are obtained.

TABLE I. PIANO MONOTONE CLASSIFICATION RESULTS

	The method of this article	The method of reference [4]
1	79.51	77.86
2	88.75	72.33
3	80.99	73.95
4	83.72	74.99
5	84.06	71.84
6	87.76	71.30
7	82.92	70.25
8	86.26	68.16
9	78.73	68.42
10	80.32	73.31
11	88.74	73.88
12	83.59	76.15

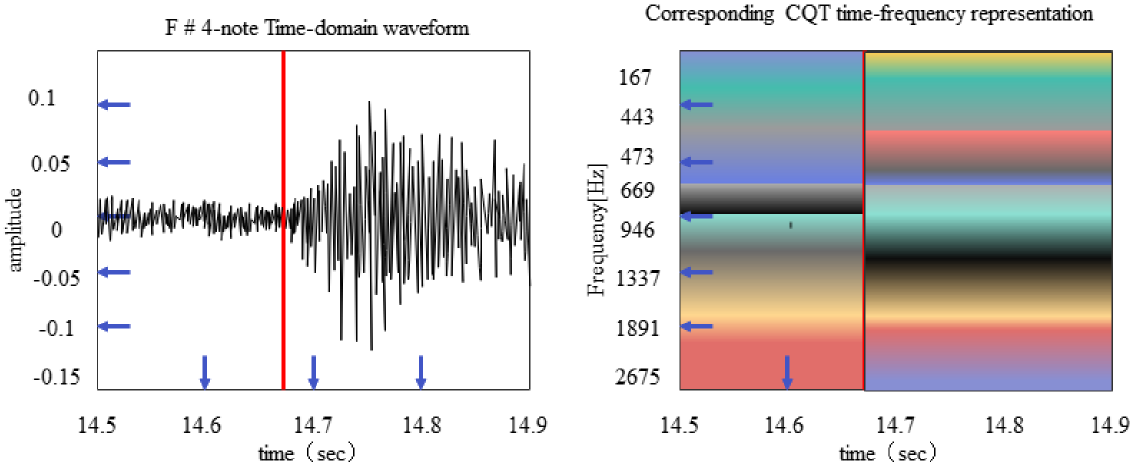


Fig. 6. F#4 note time domain diagram and corresponding CQT spectrum diagram.

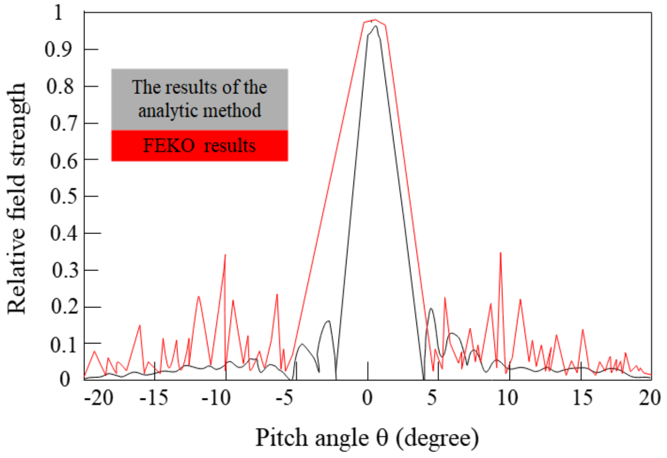


Fig. 7. The normalized direction diagram of the parabolic piano tone transmission hardware simulated in FEKO.

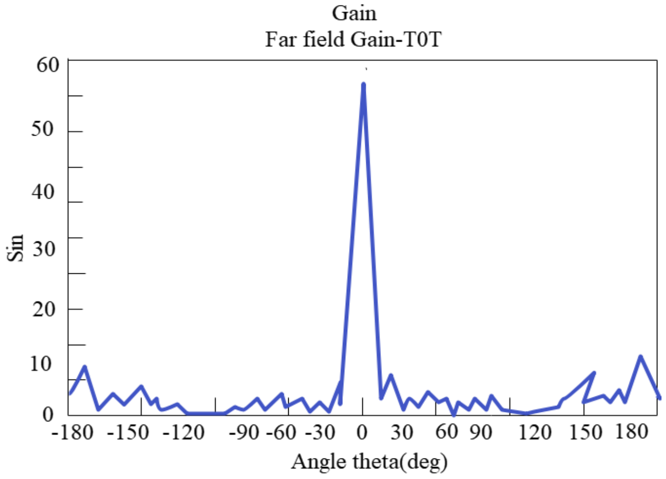


Fig. 8. Gain of piano tone transmission hardware.

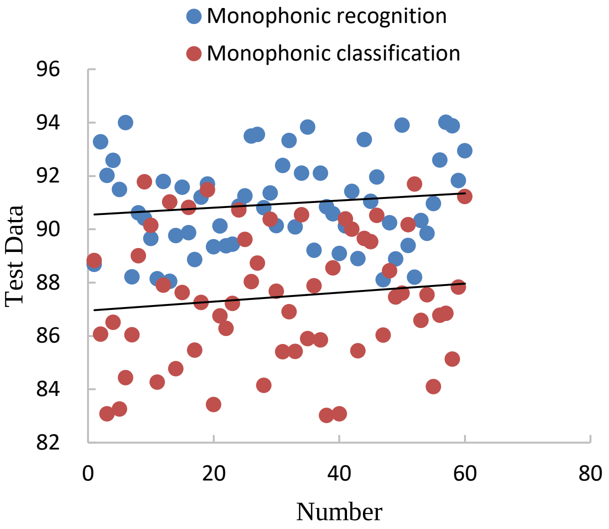


Fig. 9. Verification of the effect of the piano single tone recognition and classification method based on the CNN model.

C. Analysis and Discussion

Nowadays, piano music on the internet generally stores digital audio information, which is generated by sampling analog audio data. The higher the sampling frequency, the larger the amount of digital audio data, and the better the fidelity of digital audio. Digital audio, as a manifestation of acoustic signals, can also be analyzed in the frequency domain and decomposed into pitch components of different frequencies. The frequency domain is composed of many sine functions (or cosine functions, or a combination of both), which have different amplitudes and phases, representing the rich information of audio in the frequency domain angle. The digital audio information stored in piano music is directly unfolded as the energy amplitude trend that changes over time, which is called the time-domain characteristics of audio signals and is the most intuitive representation of the signal. Time domain analysis and frequency domain analysis consider signal characteristics from two perspectives. For digital audio, the amount of data in the time domain is very large. The higher the sampling frequency of audio, the larger the amount of data, which often leads to a large computational workload. Compared to time-domain analysis, frequency-domain analysis has a smaller amount of data and can better reflect some substantive features. Therefore, frequency domain analysis has gradually become the mainstream of signal analysis. Frequency domain features are the results of frequency domain analysis, which describe the basic characteristics of audio signals. Using frequency domain features to represent music is not only easy to implement, but also reduces data volume and facilitates data processing.

FEKO has an algorithm that combines the method of moments as well as the more classical high-frequency analysis methods (physical optics) and consistent diffraction theory. It is extremely suitable for application in analyzing the design layout of piano tone transmission hardware, radar cross section (RCS) and other electromagnetic field analysis problems.

For directional piano tone transmission hardware such as parabolic piano tone transmission hardware, the main consideration is the main lobe pattern and gain of the piano tone transmission hardware. Therefore, the calculation results in this paper can be applied to engineering calculations.

From the results of piano single tone classification, the quantitative evaluation results of the method proposed in this article are distributed between [78, 89], while the quantitative evaluation results of the method in study [4] are distributed between [68, 8]. Therefore, it can be seen that the method proposed in this article has certain advantages compared to traditional methods

From the above research, it can be seen that the piano single tone recognition and classification method based on CNN model proposed in this paper can play an important role in piano single tone recognition.

V. Conclusion

Music and speech are both sound signals, and the basic principles of their recognition are similar. That is to say, all of the sound signals are analyzed, and processing processes such as noise processing, feature analysis, and recognition must be applied. The single-tone signal of the piano, as a sound signal, follows the basic laws of acoustics. In the field of speech research, the current technology is relatively stable and mature, so this paper draws on and refers to the technology of speech recognition and applies it to single-speech recognition. At the same time, there are obvious differences between speech and music: individual differences in speech are large. Even if the same person says the same sentence twice, the sound is quite different and cannot be exactly the same. However, for the same musical instrument, such as the piano studied in this paper, the same key is pressed by anyone at any time, the sound difference is very small, and it has a high degree of acoustic similarity. This paper combines the CNN model to build a piano single tone recognition and classification model to improve the efficiency of piano tone intelligent processing. The research shows that the piano single tone recognition and classification method based on the CNN model proposed in this paper can play an important role in piano tone recognition. The follow-up work of this article is as follows: In terms of note feature extraction, it is necessary to further address the interference of harmonic waves caused by fast rhythms. In the design of convolutional neural networks, further in-depth research is needed on the impact of network structure and loss function design on classification performance.

References

[1] Wang, X. (2018). Research on the improved method of fundamental frequency extraction for music automatic recognition of piano music. Journal of Intelligent & Fuzzy Systems, 35(3), 2777-2783.

[2] Shuo, C., & Xiao, C. (2019). The construction of internet+ piano intelligent network teaching system model. Journal of Intelligent & Fuzzy Systems, 37(5), 5819-5827.

[3] Liu, M., & Huang, J. (2021). Piano playing teaching system based on artificial intelligence–design and research. Journal of Intelligent & Fuzzy Systems, 40(2), 3525-3533.

[4] Lampe, R., Turova, V., & Alves-Pinto, A. (2019). Piano jacket for perceiving and playing music for patients with cerebral palsy. Disability and Rehabilitation: Assistive Technology, 14(3), 221-225.

[5] Johnson, D., Damian, D., & Tzanetakis, G. (2020). Detecting hand posture in piano playing using depth data. Computer Music Journal, 43(1), 59-78.

[6] Gruhn, W., Ristm, R., Schneider, P., D'Souza, A., & Kiilu, K. (2018). How stable is pitch labeling accuracy in absolute pitch possessors? Empirical Musicology Review, 13(3-4), 110-123.

[7] Schutz, M., & Gillard, J. (2020). On the generalization of tones: A detailed exploration of non-speech auditory perception stimuli. Scientific reports, 10(1), 1-14.

[8] Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., ... & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. Proceedings of the National Academy of Sciences, 115(28), E6630-E6639.

[9] Dean, R. T. (2022). The Multi-Tuned Piano: Keyboard Music without a Tuning System. Leonardo, 55(2), 166-169.

[10] Mahanta, S. K., Khilji, A. F. U. R., & Pakray, P. (2021). Deep Neural Network for Musical Instrument Recognition Using MFCCs. Computación y Sistemas, 25(2), 351-360.

[11] Hong, Y., Chau, C. J., & Horner, A. (2018). How often and why mode fails to predict mood in low-arousal classical piano music. Journal of New Music Research, 47(5), 462-475.

[12] Bando, Y., & Tanaka, M. (2022). A Chord Recognition Method of Guitar Sound Using Its Constituent Tone Information. IEEJ Transactions on Electrical and Electronic Engineering, 17(1), 103-109.

[13] Jiam, N. T., & Limb, C. (2020). Music perception and training for pediatric cochlear implant users. Expert Review of Medical Devices, 17(11), 1193-1206.

[14] Foley, L., & Schutz, M. (2021). High time for temporal variation: Improving sonic interaction with auditory interfaces. IEEE Instrumentation & Measurement Magazine, 24(7), 4-9.

[15] Chari, D. A., Barrett, K. C., Patel, A. D., Colgrove, T. R., Jiradejvong, P., Jacobs, L. Y., & Limb, C. J. (2020). Impact of auditory-motor musical training on melodic pattern recognition in cochlear implant users. Otology & Neurotology, 41(4), e422-e431.

[16] Dissegna, D., Sponza, M., Falleti, E., Fabris, C., Vit, A., Angeli, P., ... & Toniutto, P. (2019). Morbidity and mortality after transjugular intrahepatic portosystemic shunt placement in patients with cirrhosis. European Journal of Gastroenterology & Hepatology, 31(5), 626-632.

[17] Pamies, S. (2021). Deconstructing modal jazz piano techniques: The relation between Debussy's piano works and the innovations of post-bop pianists. Jazz Education in Research and Practice, 2(1), 76-105.

[18] Miyazaki, K. I., Rakowski, A., Makomaska, S., Jiang, C., Tsuzaki, M., Oxenham, A. J., ... & Lipscomb, S. D. (2018). Absolute pitch and relative pitch in music students in the East and the West: Implications for aural-skills education. Music Perception: An Interdisciplinary Journal, 36(2), 135-155.

Intelligent Evaluation and Optimization of Postgraduate Education Comprehensive Ability Training under the Mode of “One Case, Three Systems”

Yong Xiang[1], Zeyou Chen[2], Liyu Lu[3], Yao Wei[4]*
School of Civil Engineering, Architecture and Environment, Xihua University, Chengdu, Sichuan, 610039, China[1, 2, 3]
School of Construction Engineering, Sichuan Technology and Business University, Chengdu, Sichuan, 611745, China[4]

***Abstract*—This study aims to explore the intelligent evaluation and optimization methods for the comprehensive ability training of graduate students under the mode of "one case, three systems" to improve the quality and effect of graduate training. Firstly, a weighted clustering algorithm for mixed attributes is designed. Secondly, an evaluation model of postgraduate training quality based on sampling method and ensemble learning is established. Finally, the algorithm's and the model's performance are compared and tested. The test results show that with the increase in the number of experiments, the accuracy of the proposed weighted clustering algorithm can reach more than 90%, which is improved by 10%. The average number of iterations is 276, and the accuracy and F1 value can achieve the highest level with fewer iterations and stable algorithm performance. Compared with the R1 model, F1 and the accuracy of the model proposed in this study are enhanced by 3.29% and 6.75%, respectively. The feature-weighted clustering algorithm and the training quality evaluation model designed here complement each other and jointly construct a more elaborate and comprehensive training system. The feature-weighted clustering algorithm oriented to mixed attributes for the first time combines sampling methods and ensemble learning in the education ability training. Moreover, a multi-dimensional and intelligent postgraduate training evaluation framework is constructed, which provides a new idea for improving the quality of postgraduate training.**

Keywords*—Comprehensive ability training of graduate students; one case; three systems; feature weighted clustering algorithm; sampling method; ensemble learning

I. Introduction

With society's continuous development and change, higher education has faced increasingly complex and diverse challenges in recent years. Improving the quality and comprehensive ability of postgraduate education has become one of the key objectives of the reform and development of higher education [1], [2]. However, the traditional postgraduate training mode has gradually revealed its limitations in meeting the growing social needs and personnel training requirements. Higher education in China is actively exploring and promoting the reform of the education system [3], [4].

In higher education, especially in postgraduate education, intelligent evaluation, and optimization research has gradually attracted widespread attention. Jiang (2021) utilized data analysis and mining technology to comprehensively evaluate the academic achievements, scientific research activities, and practical experience of graduate students to understand the characteristics and potential of students better and provide a basis for personalized training programs [5]. He (2020) discussed the critical issues of cultivating the comprehensive ability of graduate students. For example, how to reasonably integrate elements such as case teaching, curriculum system, mentor system, and practical system and establish effective connections and interactions between these elements to achieve the best effect of comprehensive ability training [6]. Liu (2023) focused on how to use big data analysis (BDA), machine learning (ML), and artificial intelligence (AI) technology to quantitatively evaluate the academic performance, innovation ability, and practical experience of graduate students [7].

The importance of this study cannot be ignored, as it deeply explores intelligent evaluation and optimization methods in the cultivation of postgraduate comprehensive abilities. In the current field of education, although research on intelligent evaluation and personalized learning is increasing, how to effectively integrate these concepts and technologies into postgraduate education models, especially in the application of the "one case, three systems" model, is still an area that has not been fully explored. In addition, the practical application of music education and virtual reality (VR) technology in the education system and how these technologies affect the quality and effectiveness of postgraduate ability development also require more systematic research and evaluation. Hence, this study aims to fill the gap in existing research by introducing intelligent technology and advanced algorithms, enriching understanding and comprehension of intelligent evaluation, personalized learning paths, music education integration, and VR technology application in higher education. Through the implementation of this study, a new postgraduate training evaluation framework based on data-driven and intelligent analysis can be constructed. Moreover, the effectiveness and advantages of intelligent evaluation and optimization methods in practical applications can be more clearly revealed through empirical research. As a result, the quality of postgraduate education has been improved while providing educators with more accurate training decision support. Additionally, through the in-depth analysis and empirical results provided by this study, it is possible to better understand the design principles of personalized learning pathways and how music education and VR technology can be creatively integrated into modern education systems. It can offer new theoretical and practical perspectives for future research in related fields. Compared to previous studies, this study adopts an advanced weighted clustering algorithm for mixed attributes and the evaluation model based on sampling method and ensemble learning, a significant extension of existing educational evaluation methods. The limitation of this study is that the effectiveness of intelligent evaluation and optimization methods may be limited by factors such as data quality, algorithm adaptability, and diversity of teaching environments. Future research must validate the universality and stability of the proposed model in a broader educational context, thus enhancing comparative fairness and the universality of conclusions. Through continuous iteration and optimization, this study aims to improve the intelligent evaluation and optimization framework for cultivating graduate comprehensive abilities, making it an indispensable part of a high-quality postgraduate training system.

In short, with the rapid development of information technology, many colleges and universities began to explore the use of AI, data mining (DM), ML, and other technical means to achieve the evaluation of graduate students' comprehensive ability and the optimization of training programs. Section I summarizes the research background, current situation, and purpose. Section II proposes a weighted clustering algorithm for mixed attributes to solve the attribute evaluation problem. Then, an evaluation model of postgraduate training quality based on sampling method and ensemble learning is established. Section III tests the performance of the algorithm and model proposed here, describes the test data, environment, and results, and compares and discusses the proposed method with the traditional method. Section IV looks forward to the research contribution and future direction, explaining this study's theoretical and practical value and development space. Section V concludes the study.

II. Literature Review

In the research field of "One Case, Three Systems," under the intelligent evaluation and optimization of postgraduate education comprehensive ability training mode, the work of multiple scholars provides key background and support for the current research. Li et al. (2020) introduced the entropy weight method and grey clustering analysis, providing a novel method for evaluating the quality of online teaching [8]. This reflected the scholars' understanding of using multi-dimensional evaluation when facing the problems of online teaching quality. Based on the proposed evaluation model, scholars also proposed a series of strategies to improve the quality of online teaching, thus enriching the practical application of the research. However, this study had limitations, such as the representativeness and feasibility of the datasets used in empirical analysis. Future research can further validate and expand the conclusions of this study through broader data collection and deeper empirical research. Lee et al. (2021) focused on a holistic music education approach for children jointly developed by music therapists and experts, combining technology and music, integrating local culture, and constructing a holistic education framework [9]. The research results indicated that implementing a holistic music education approach can significantly enhance children's abilities with developmental delays, while supportive training has a positive effect. In addition, decision trees explored and developed an intelligent evaluation model with high learning effectiveness. The sensitivity of this model within the sample reached 90.6%, while the comprehensive indicator F was 79.9%, which had a high reference value. In the future, educators can use DM to assist decision-making systems as evaluation tools to evaluate children participating in education in the early and middle stages of the curriculum. It can also predict their continuous implementation and learning effectiveness, help decide whether to continue investing and adjusting the curriculum, and use educational resources more effectively. Yang et al. (2022) investigated the evolution from Assessment of Learning (AoL) to Assessment as Learning (AaL) from four aspects: participants, testing format, process-based multivariate data, and measurement models for multivariate data. They proposed suggestions for interdisciplinary collaboration, integrating education, psychology, and information technology into the theories and methods of educational measurement [10]. In addition, the study emphasized that measurements' effectiveness, ethics, and fairness should also be considered vital issues. Researchers and practitioners in educational measurement must adhere to the pursuit of the substantive significance of measurement and provide unique experience and guidance for the theoretical and practical development of educational evaluation in this tremendous revolution. Wang et al. (2022) constructed a personalized learning model based on distributed computing methods of the Internet of Things (IoT) and clustering algorithms of deep learning (DL). The research results showed that the accuracy of this model on personalized learning platforms based on the IoT and DL algorithms reached 85% [11]. Compared with the latest research models, this model performed better in score prediction and customized recommendations. This model had significant practical value in promoting the development of the IoT and DL in professional learning. This exploration innovatively divided the understanding level of learners at a hierarchical level. It provided personalized learning resources aligned with their cognitive abilities to enhance their knowledge level and achieve personalized learning. These studies focused on improving algorithm performance, offering relevant research for the performance improvement of weighted clustering algorithms in this study. Lee and Hwang (2022) highlighted sustainable education from the perspective of VR technology application, providing innovative insights into the educational research field [12]. By delving into the experience of pre-service teachers in VR content design, they highlighted the importance of this transformative experience in enhancing teachers' technical readiness, digital citizenship, and perceived educational benefits. The study emphasized the importance of introducing emerging technologies into education to promote sustainable education development.

Although the research of the above scholars provided valuable insights into the current fields of intelligent evaluation, personalized learning, music education, and the application of VR technology in education, there are certain limitations to each research. Firstly, many studies were influenced by specific backgrounds, sample sizes, and dataset limitations, which limited their universality and generalization ability. Secondly, the empirical analysis of some studies lacked sufficient breadth and depth, failing to validate the applicability of the proposed model. Besides, some studies still lacked practical application cases in real-world environments, thus limiting their guiding role in practice. However, examining previous research provided this study with a profound understanding of different aspects of the field of education. Still, these findings must be validated in a broader and more diverse context. In addition, there was a relative lack of comprehensive research on cultivating graduate students' extensive abilities under the "One Case, Three Systems" model. Previous research aimed to fill some gaps in previous studies and expand understanding of the practical applications of personalized learning, intelligent evaluation, music education, and VR technology in education. By delving deeper into the limitations of previous research, there was an urgent need for broader, deeper, and more empirical research to enhance the practicality of existing theoretical frameworks and promote their practical application. This study's importance was applying the feature-weighted clustering algorithm and training quality evaluation model to build a more comprehensive framework for postgraduate training evaluation. Compared with the previous research, this study emphasized the total consideration of multi-dimensional characteristics while improving the evaluation model's algorithm performance and accuracy.

III. Method

A. Feature-weighted Clustering Algorithm for Mixed Attributes

The feature-weighted clustering algorithm for mixed attributes proposed in this study is a data clustering technology that deals with numerical and discrete (mixed attributes) data [13,14]. The flow of the clustering algorithm is displayed in Fig. 1:

In Fig. 1, the proposed algorithm's core idea is to give appropriate weights to different attributes in the clustering process to accurately reflect the importance of attributes when calculating the distance or similarity between data points [15]. The specific process is as follows:

Step 1: Improvement of dissimilarity of classification attributes;

Sample distribution information is described by covariance matrix, and the Mahala Nobis distance D_m^2 between sample x_i and cluster center a_j is defined by Eq. (1):

$$D_m^2 = (x_i - a_j)\Sigma^{-1}(x_i - a_j)^T \tag{1}$$

Σ represents covariance matrix, which Eq. (2) can calculate:

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} x_i {x_i}^T - a_j {a_j}^T \tag{2}$$

On this basis, the proportional coefficient P is introduced to increase the statistical probability of classified samples and unclassified samples in clustering, as illustrated in Eq. (3):

$$P = diag\left(\sqrt{\gamma_1}, \sqrt{\gamma_2}, \cdots, \sqrt{\gamma_m}\right), \gamma = 1 - |C_l| \,/\, |C_{lij}| \tag{3}$$

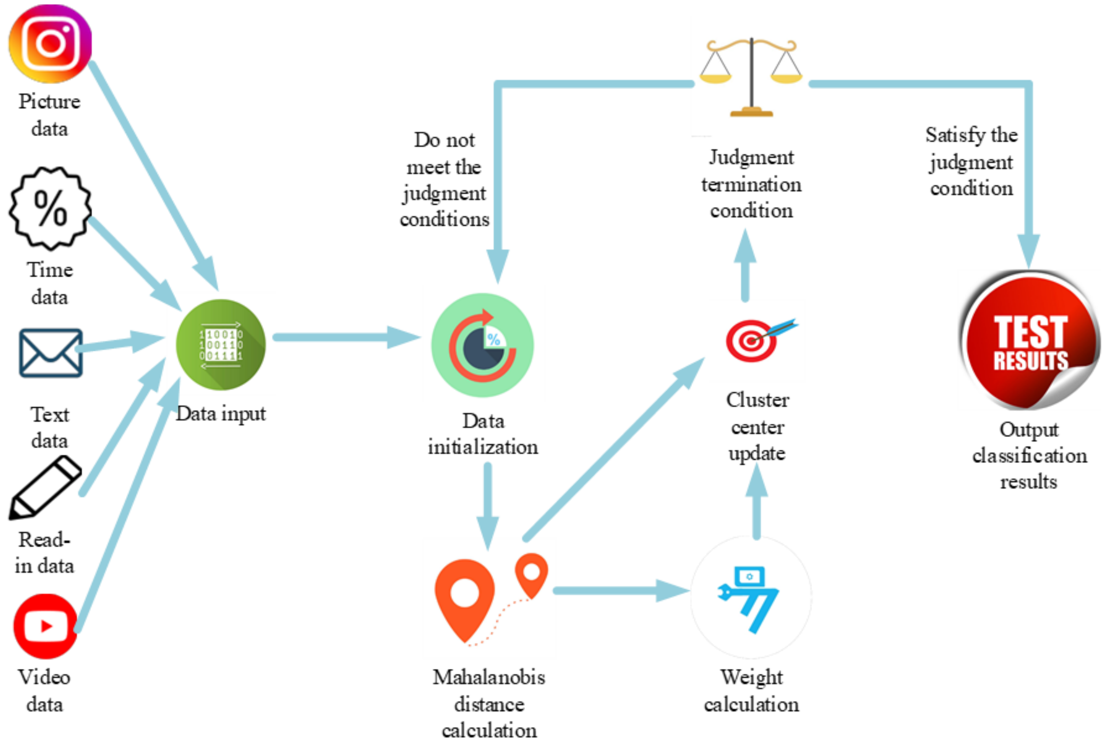

Picture data
Time data
Text data
Read-in data
Video data
Data input
Do not meet the judgment conditions
Data initialization
Mahalanobis distance calculation
Judgment termination condition
Cluster center update
Weight calculation
Satisfy the judgment condition
TEST RESULTS
Output classification results

Fig. 1. Flow chart of feature-weighted clustering algorithm for mixed attributes.

$|C_l|$ refers to the number of classified samples, and $|C_{lij}|$ denotes the frequency of the *j*-th attribute of unclassified sample x_i in C_l . Redefine Mahalanobis distance after orthogonal decomposition of positive definite matrix Σ, as listed in Eq. (4):

$$D_m^2 = \left[PQ\Lambda^{-\frac{1}{2}}Q^T(x_i - a_j)\right]^T \left[PQ\Lambda^{-\frac{1}{2}}Q^T(x_i - a_j)\right] \quad (4)$$

Q is the inverse of the covariance matrix, the off-diagonal element of Λ is 0, and the diagonal element is the eigenvalue of the covariance matrix.

Step 2: Calculation of the weights of the numerical attributes;

The *K* nearest neighbor sample of x_i reads:

$$KN(x_i) = \{x_j | d(x_i, x_j) \le d(x_i, Near(x_i))\} \quad (5)$$

$Near(x_i)$ represents the nearest *K*-th sample point to the Mahala Nobis of x_i. The definition of intra-cluster dispersion d_1 is as follows:

$$d_1 = \sum_{j=1}^{k} \frac{|x_{ij} - KN(x_{ij})|}{max(A_l) - min(A_l)} \quad (6)$$

A_l represents the cluster's feature set of sample points. The definition of cluster dispersion is expressed by Eq. (7):

$$d_2 = \sum_{C} \frac{p(C)}{1 - p(x_i)} \sum_{j=1}^{k} \frac{|x_{ij} - KN(x_{kj})|}{max(A_l) - min(A_l)} \quad (7)$$

$C \neq class(x_i)$. The update of the weights of samples and *X* on the feature set A_l is Eq. (8):

$$w_l^n = w_l^n - \frac{d_1}{mk} + \frac{d_2}{mk} \quad (8)$$

The weights of numerical attributes are mainly determined by d_1 and d_2. The smaller the intra-cluster dispersion is, the greater the inter-cluster dispersion is, and the higher the similarity of samples in changing attributes. On the contrary, the lower the discrimination of attributes for clustering, the greater the weight.

Step 3: Calculation of the weight of the classification attribute [16,17];

$D(A_l) = \{a_l | a_l = x_{il}, 1 \le i \le n, P \le l \le m\}$ is defined as the set consisting of the first attribute of the sample set *X*, and its distribution function is as follows:

$$\begin{cases} p(a) = \frac{|D(A_{l=v})|}{|D(A_l)|} \\ p(r) = \frac{|D(R_{l=t})|}{|D(R)|} \\ p(a,r) = \frac{|D(A_{l=v})|}{|D(R)|} \end{cases} \quad (9)$$

$p(a)$ and $p(r)$ are the marginal probability distribution function of classification attribute *A* and clustering result R; $p(a, r)$ is the joint probability distribution function of *A* and *R*. The mutual information *M* between *A* and *R* can be written as Eq. (10):

$$M(A_l, R) = \sum_{r \in R} \sum_{a \in A_l} p(a, r)\, log\, 2\, \left(\frac{p(a,\ r)}{p(a) * p(r)}\right) \quad (10)$$

The weight of the classification attribute is calculated as Eq. (11):

$$w_l^C = \frac{M(A_l, R)}{\sum_{l=p+1}^{m} M(A_l, R)} \quad (11)$$

Based on these calculations, the revised objective function $E(U, C)$ is obtained, as indicated in Eq. (12):

$$E(U, C) = \sum_{j=1}^{k} \sum_{i=1}^{n} u_{ij}{}^{a}(w_l^n d_1(x_i - a_j) + \gamma w_l^C d_2(x_i - a_j)) \quad (12)$$

Step 4: Update of the cluster center;

The updating equation of the l-th numerical feature c_{jl} of the cluster center c_j is:

$$c_{jl} = \frac{\sum_{i=1}^{n} (u_{ij})^a w_l^n x_{il}}{\sum_{i=1}^{n} (u_{ij})^a} \quad (13)$$

The equation for updating the l-th classification feature c_{il} of the cluster center c_i reads:

$$c_{il} = a_l^s \in D(A_l) \quad (14)$$

s satisfies the condition shown in Eq. (15):

$$\sum_{i=1}^{n} (u_{ij}{}^{a} | x_{ij} = a_l^s) \ge \sum_{i=1}^{n} (u_{ij}{}^{a} | x_{ij} = a_l^t), s \ge 1, t \le |D(A_l)| \quad (15)$$

$|D(A_l)|$ is the number of values of the classification attribute A_l.

Step 5: Initialization of parameters, cluster center matrix, and iteration times;

All variables start from 0, and the maximum number of iterations is set to t.

Step 6: The weighted clustering is completed according to the process shown in Fig. 1.

B. *Evaluation Model of Postgraduate Training Quality based on Sampling Method and Ensemble Learning*

The evaluation model of postgraduate training quality based on sampling method and ensemble learning is an analytical tool for evaluating the process and effect of postgraduate training [18]. The specific structure is presented in Fig. 2.

Fig. 2 signifies that postgraduate training involves data of multiple dimensions, such as academic achievements, scientific research activities, and practical experience. To reduce computational complexity and improve efficiency, the model extracts feature from different training dimensions and gives appropriate weights to different features in the selected samples, accurately reflecting their importance in training quality evaluation.

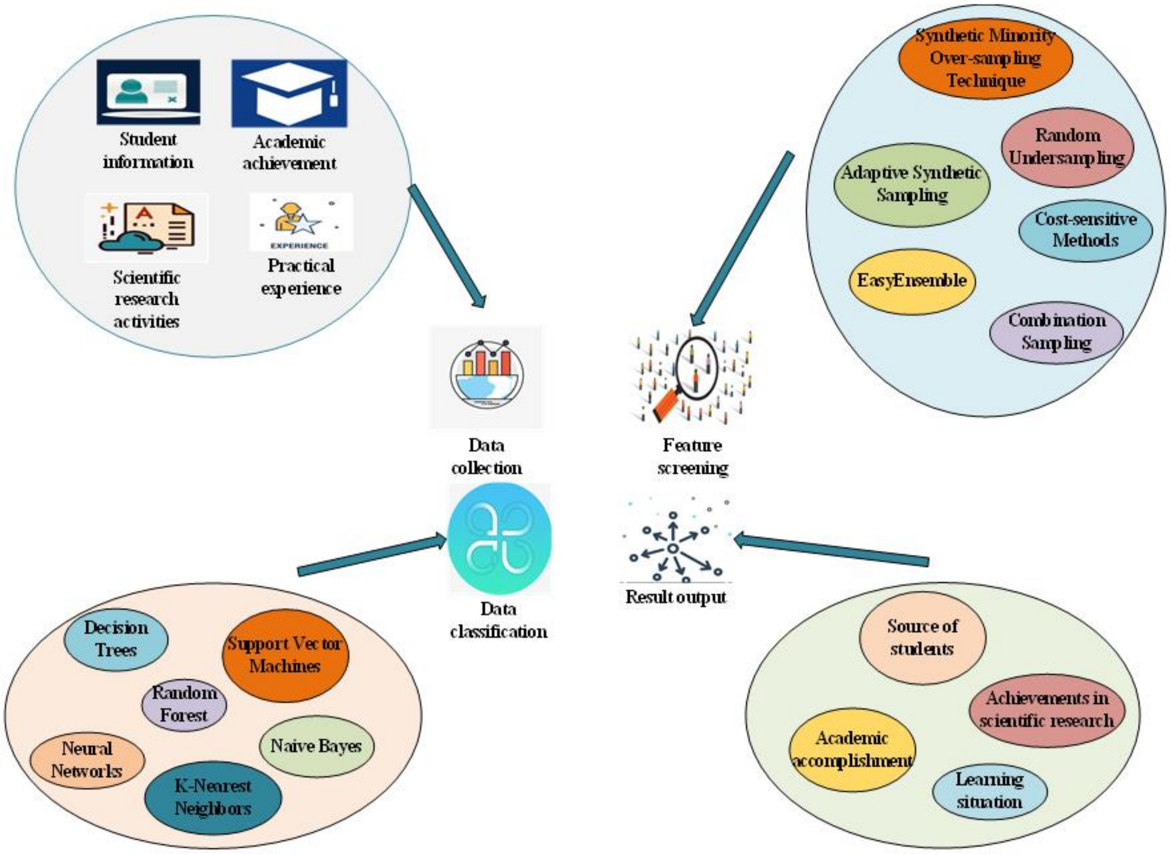


Fig. 2. Model structure of postgraduate training quality evaluation based on sampling method and ensemble learning.

IV. Results and Discussion

A. Data Collection and Experimental Environment

This study employs the Student Performance Data Set data from the University of California, Irvine's UCI ML Repository database as the algorithm performance test's test set and data source. The UCI data set is an open-source dataset proposed by the University of California, Irvine, which is suitable for pattern recognition and ML direction. Many scholars choose to use the data set on UCI to verify the correctness of their proposed algorithms. These data sets are divided into binary classification, multi-classification, and regression fitting problems. The UCI data set provides the main attributes of each data set. It can be used to demonstrate the rationality of various algorithms proposed by oneself through experimental results on its data set. Data address: https://archive.ics.uci.edu/datasets. The Student Performance Data Set aims to predict the performance of students in two secondary schools in Portugal, covering two different subjects: Mathematics (mat) and Portuguese (por). The data collection methods include school reports and questionnaire surveys, encompassing multiple attributes such as student performance, demographics, and social and school-related characteristics [19]. The entire data set contains 649 instances, mainly used for classification and regression tasks. Two independent data sets correspond to two subjects: mat and por. In both data sets, the target attribute G3 refers to the final year grade, while G1 and G2 represent the first and second-semester grades, respectively. It is important to note that there is a strong correlation among G3, G2, and G1. This is because G3 is the final grade (released in Semester 3), while G1 and G2 correspond to Semester 1 and 2 grades. Although it is more challenging to predict G3 without G2 and G1, this prediction is more useful in practical applications. The attributes of this data set include school, gender, age, place of residence, family size, parental cohabitation status, parental education levels, parental occupation, the reason for choosing a school, guardians, school time, study time, number of past classroom failures, additional educational support, family educational support, extra paid courses, and extracurricular activities. Additionally, it also involves whether to attend daycare, whether to plan for higher education, whether there is a romantic relationship, family network access, quality of family relationships, free time after school, frequency of outings with friends, alcohol consumption on workdays and weekends, current health status, and school absences. Data address: https://archive.ics.uci.edu/dataset/320/student+performance.
This study tests the evaluation model of postgraduate training quality based on sampling method and ensemble learning. The ratio of the training set to the data set is 4:1, and the number of training set samples is 622, divided into three levels: high, medium, and low, with the numbers 68, 480, and 74, respectively. The test set's data volume is 156, and the number of high, medium, and low samples is 17, 121, and 18, respectively, which meets the proportional requirements.

The experimental equipment environment of this study is Dell notebook equipment; the CPU model is Intel (R) Core

(TM) i5-7300HQ CPU @ 2.50GHz; the memory is 8G; the graphics card model is GTX1050; the operating system is Windows 11 system, and the software environment is C++ language environment to ensure the stability of the algorithm and model in the process of training and testing.

B. Indicator Setting

1) Feature-weighted clustering algorithm for mixed attributes: This study uses the accuracy F-Measure to evaluate the algorithm's performance. F-Measure is the weighted harmonic average of accuracy and recall. Accuracy indicates the proportion that objects in the same cluster are divided into the same category, and recall reflects the proportion that objects in the same category are assigned to the same cluster. The fuzzy factor is set to 2 in the experiment, the clustering threshold is 0.0001, and the maximum number of iterations T is set to 10000.

When the algorithm is used to verify the comprehensive ability training of postgraduate education, it is verified by the evaluation results of the postgraduate training information data of S University from 2019 to 2022 according to the indicators exhibited in Table I:

Table I describes that 11 secondary refinement indicators are set under the classification of four primary indicators: student source, tutor title, scientific research achievements, and learning situation, and weighted cluster analysis is carried out.

a.

2) Evaluation model of postgraduate training quality based on sampling method and ensemble learning: Aiming at the designed evaluation model, this study evaluates it through three standards: accuracy, recall, and F1 value. F1 value, as a calculation index of the reconciliation of accuracy and recall, can fully reflect the classification ability of the model and is the most representative.

C. Experimental Result

To verify the accuracy of the proposed feature-weighted clustering algorithm, this algorithm is compared with the fuzzy clustering, the goal-Okumoto-kapur with k-prototypes, synthetic minority over-sampling technique, the Adaptive Synthetic Sampling, and the improved genetic fuzzy K-Prototypes algorithms. They are named A1, B2, C3, D4, E5, F6 in turn. Fig. 3 depicts various algorithms' clustering accuracy and clustering effect on data sets.

In Fig. 3(a), as the number of experiments increases, the proposed weighted clustering algorithm can achieve over 90%, which is 10% higher. Fig. 3(b) denotes that the average number of iterations of the algorithm proposed here is 276, and the accuracy and F value can reach the highest level with fewer iterations and stable algorithm performance. Fig. 4 demonstrates the proposed algorithm’s weighted clustering results for postgraduate training quality and various algorithms’ contour change curves with the change of fuzzy factors.

TABLE I. QUALITY TRAINING INDEX OF POSTGRADUATE EDUCATION

First Index	**Source of students**			
Second index	Undergraduate colleges A#	Learning mode B#	Entrance examination marks C*	
First Index	**Academic literacy**			
Second index	Tutor title D#		Tutor academic achievements E*	
First Index	**Achievements in scientific research**			
Second index	Number of papers F*	Paper quality #G*	Number of patents H*	Number of awards I
First Index	**Learning conditions**			
Second index	Average scores J*		Type of scholarships K#	

Note: * indicator is a numerical indicator, # indicator is a classified indicator.

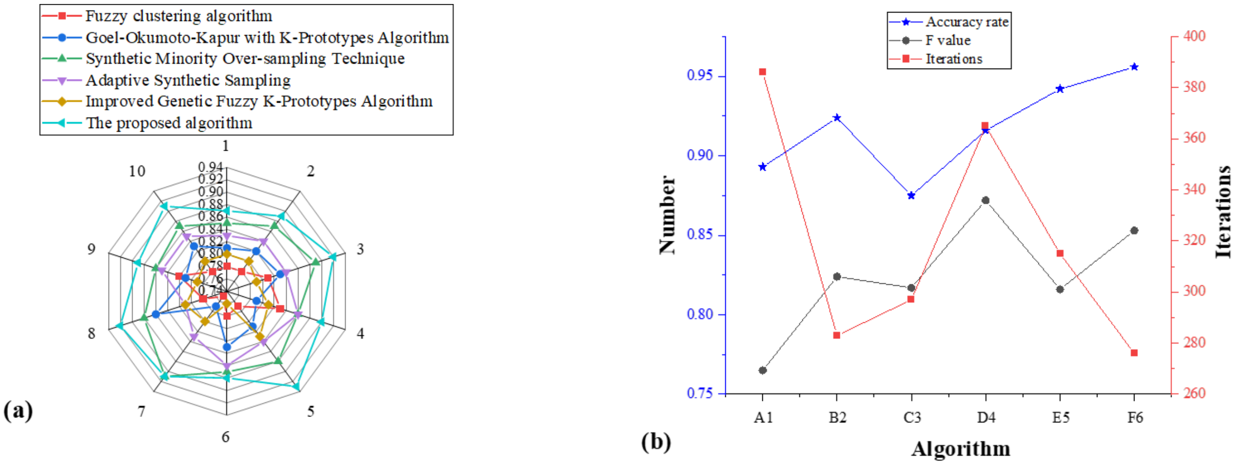


Fig. 3. Clustering accuracy and clustering effect of different algorithms on data sets ((a) Clustering accuracy; (b) Clustering effect).

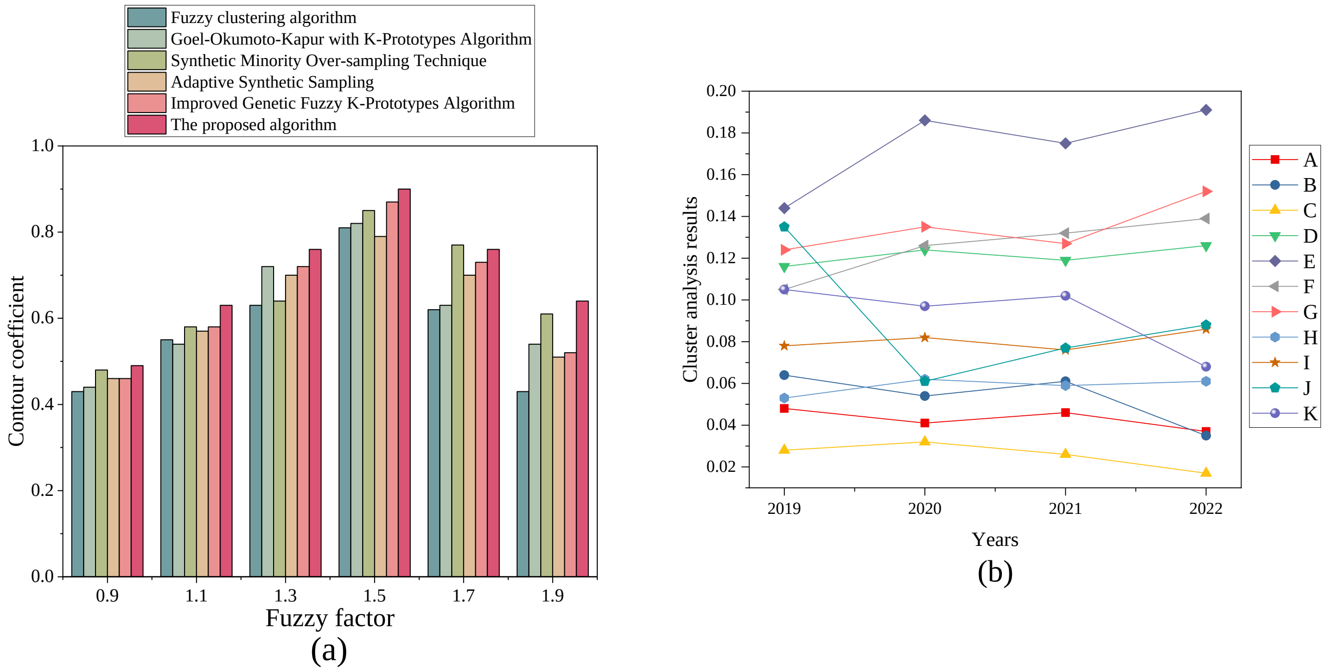


Fig. 4. Evaluation results of postgraduate training quality (a) Changes of contour coefficients of various algorithms; (b) Weighting cluster analysis results for evaluation elements)

In Fig. 4(a), with the fuzzy factor gradually increasing from 0.8 to 2.0, the contour coefficient of the proposed algorithm is higher than other algorithms, and the average value is around 0.2. It illustrates that the weighted clustering algorithm designed in this study can guarantee the data features to the greatest extent and has great advantages for feature extraction. In Fig. 4(b), the weight of evaluation indicators has not changed significantly in recent years. The academic achievements of tutors and the quality of papers published by students have become the most influential factors in the cultivation of postgraduate education ability, accounting for a relatively high proportion, with a comprehensive rate of over 46%. It can be found that the guidance of mentors to students and the students' research results are the core content of current postgraduate education ability cultivation. Fig. 5 presents the comparison of diverse classification and sampling algorithms' evaluation performance:

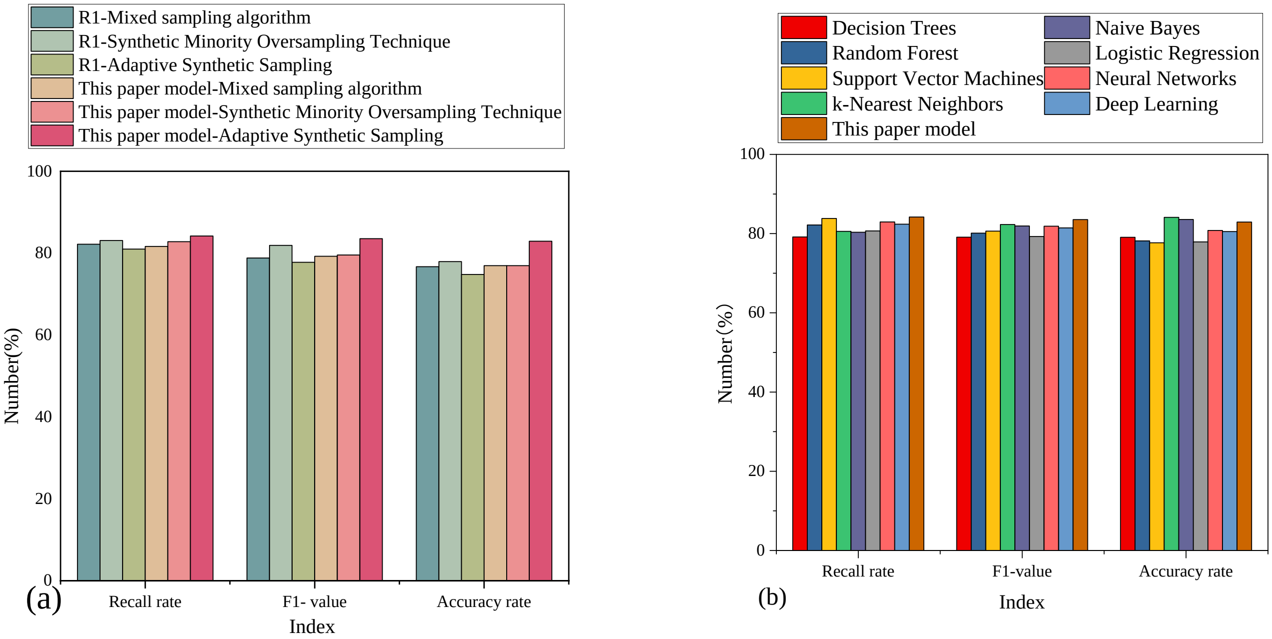


Fig. 5. Comparison of evaluation performance of different sampling and classification algorithms ((a) Comparison of sampling algorithms; (b) Performance comparison of classification algorithms).

In Fig. 5, this study sets the sampling comparison algorithm as a mixed sampling algorithm, synthetic minority oversampling technique algorithm, and Adaptive Synthetic Sampling algorithm. It constructs the R1 model based on a decision tree and random forest as base classifiers. As a meta-classifier, the Gradient Boosting Decision Tree is a quality evaluation model of postgraduate education based on sampling method and ensemble learning. Then, the model's performance is tested. In Fig. 5(a), the F1 value of the model proposed in this study is 83.54%. The accuracy is 82.91%, and its practical application performance is good. Compared with the R1 model, the proposed model’s F1 value and accuracy increase by 3.29% and 6.75%, respectively. In Fig. 5(b), compared with other algorithms, the proposed model has a recall of 84.18%, an F1 value of 83.54%, and an accuracy of 82.91%. This model has higher accuracy and stability and can effectively solve the problems of subjectivity and data imbalance in the evaluation process.

The performance evaluation results of the feature-weighted clustering algorithm on different data sets are outlined in Table II. The EduNLP Student Performance Data Set in Table II analyzes students' literal expressions and evaluates their academic performance and needs through natural language processing technology. Student Performance in Online Learning Data Set is employed to evaluate students' performance in online learning, including learning time, interaction, and other characteristics. Student Academic Performance Data Set (Kaggle) is used to analyze students' academic performance, involving grades, subject preferences and other characteristics. The results reveal that among all data sets, the feature-weighted clustering algorithm performs best on the UCI data set, at 90%, while the performance is relatively low on the Online Learning Data Set, at 81%. On different data sets, the average number of iterations of the algorithm fluctuates in a relatively reasonable range, and the average number of iterations on UCI data set is the lowest, at 276. The algorithm on the UCI data set performs the best in terms of contour coefficients, at 0.2, while the Online Learning Data Set is at 0.16. In all data sets, feature-weighted clustering algorithms use evaluation indicators such as F1 value, accuracy, and recall, providing researchers with a comprehensive performance evaluation. These evaluation results provide a reference for applying feature-weighted clustering algorithms in different backgrounds, and the performance on the UCI data set offers potential best practices for other datasets. The performance on the EduNLP Student Performance Data Set is second, while the performance on Online Learning Data Set is relatively poor, requiring more in-depth research and improvement. The Student Academic Performance dataset performed well, approaching the performance level of the UCI data set.

The performance evaluation results of various intelligent evaluation methods on UCI data sets are suggested in Table III. The results of Table III indicate that the performance of the proposed weighted clustering algorithm for mixed attributes is the best, with a clustering accuracy of 90%, an average number of iterations of 276, and a contour coefficient of 0.2. In contrast, although the STEAM Education+Smart Greenhouse VR [20] algorithm performs relatively stable in clustering accuracy, its overall performance is poor, with a clustering accuracy of 80%, an average number of iterations of 300, and a contour coefficient of 0.15. The Analytic Hierarchy Process+Delphi method [21] achieves good clustering accuracy, reaching 85%, but slightly inferior to the proposed algorithm in other indicators. The neural fuzzy method with multiple inputs and outputs [22] has relatively low clustering accuracy and F1 value, at 78% and 73.25%, respectively. However, its average number of iterations is 350, which may have some stability issues. The binary genetic algorithm [23] and the educational data classification method of swarm intelligence [24] obtain 82% and 87% clustering accuracy, respectively. Furthermore, their performances are relatively balanced, but slightly inferior to the proposed algorithm in other indicators. In general, the proposed algorithm has excellent clustering accuracy, average number of iterations, and F1 value, proving its effectiveness in students' comprehensive ability evaluation.

TABLE II. PERFORMANCE EVALUATION RESULTS OF FEATURE WEIGHTED CLUSTERING ALGORITHMS ON DIFFERENT DATA SETS

Data Set	Cluster accuracy (%) (Mean)	Average number of iterations	Contour coefficient (Mean)	Evaluation indicators	F1 value (%)	Accuracy (%)	Recall rate (%)
Student Performance Data Set (UCI)	90	276	0.2	Feature-weighted clustering	83.54	82.91	84.18
EduNLP Student Performance Data Set	88	290	0.18	Feature-weighted clustering	82.21	83.12	81.49
Student Performance in Online Learning Data Set (Kaggle)	81	360	0.16	Feature-weighted clustering	75.12	76.05	74.76
Student Academic Performance Data Set (Kaggle)	89	280	0.22	Feature-weighted clustering	83.02	84.11	82.58

TABLE III. PERFORMANCE EVALUATION RESULTS OF DIFFERENT INTELLIGENT EVALUATION METHODS ON THE STUDENT PERFORMANCE DATA SET (UCI)

Algorithm	Cluster accuracy (%) (Mean)	Average number of iterations	Contour coefficient (Mean)	Evaluation indicators	F1 value (%)	Accuracy (%)
The proposed algorithm	90	276	0.2	83.54	82.91	84.18
STEAM Education+Smart Greenhouse VR	80	300	0.15	76.21	78.34	75.89
The Analytic Hierarchy Process+Delphi method	85	320	0.18	80.56	81.42	79.73
The neural fuzzy method with multiple inputs and outputs	78	350	0.13	73.25	75.11	72.46
The binary genetic algorithm	82	280	0.22	78.92	79.87	78.36
The educational data classification method of swarm intelligence	87	300	0.24	81.67	82.53	81.12

D. Discussion

Zhu (2023) paid attention to using intelligent methods to design personalized learning paths and course recommendation systems to improve graduate students' comprehensive ability training effect according to their interests, abilities, and backgrounds [25]. Thurzo (2023) used BDA, ML, and DM technology so researchers could analyze students' learning behavior, performance, and progress and dig out meaningful information for cultivating graduate students' comprehensive ability to evaluate and optimize their intelligence [26]. Shan (2023) focused on building a complete comprehensive ability evaluation system, including academic ability, innovation ability, practical ability, and other indicators and standards, to conduct intelligent evaluation and optimization more effectively [27]. The algorithm and model proposed in this study can achieve an accuracy of over 90% with an increase in the number of experiments, an improvement of 10%, and an average iteration of 276 times. Therefore, the accuracy and F-value can reach the highest level. With the fuzzy factor increasing from 0.8 to 2.0, the proposed algorithm's contour coefficient is larger than other algorithms, and the average value is about 0.2. The proposed model's F1 value is 83.54%, and the accuracy is 82.91%. Compared with the R1 model, the proposed model's F1 value and accuracy are improved by 3.29% and 6.75%, and the recall and F1 values are 84.18% and 83.91%. Compared with the traditional methods, the feature-weighted clustering algorithm for mixed attributes can better consider the relationship between different attributes and improve the accuracy and interpretation of clustering. The evaluation model of postgraduate training quality based on sampling method and ensemble learning can make personalized evaluation and training suggestions according to the situation of individual students to better meet their needs and development direction.

V. CONCLUSION

The proposed weighted clustering algorithm based on mixed attributes can better handle multi-attribute data and accurately evaluate graduate students' comprehensive ability. The evaluation model of graduate students' training quality based on sampling method and ensemble learning provides new ideas and tools, and personalized training suggestions can better meet the needs of graduate students. Applying the algorithm and model to emergency management helps improve the quality and effect of graduate students' training in this field. However, the proposed methods and models still need adaptive adjustment in other fields, and their versatility is limited. In future research, people can consider introducing multiple data sources, such as text, image, and voice, to carry out multimodal data fusion to evaluate graduate students' comprehensive ability. In addition, applying the proposed methods and models in other professional fields has been explored to verify their applicability and effectiveness in different fields.

ACKNOWLEDGEMENT

This work was sponsored in part by the Graduate School of Xihua University: Research on Interdisciplinary Education Model and Innovation Reform for graduate students of Emergency major based on One case and three systems (No: YJG2020008); Sichuan Department of Education: Research on the Cultivation and realization Path of National Identity Concept in College Students' Innovation and Entrepreneurship Education under the guidance of Party Building (No: JG2021-938).

COMPETING OF INTEREST

The authors declare no competing of interests.

AUTHORSHIP CONTRIBUTION STATEMENT

Yao Wei: Conceptualization, Investigation, Methodology, Writing, Project administration.

Yong Xiang: Formal analysis, Language review

Zeyou Chen: Software, Methodology

Liyu Lu: Validation

REFERENCES

[1] Y. Cao, J. Shan, Z. Gong, J. Kuang, and Y. Gao, "Status and challenges of public health emergency management in China related to COVID-19," Front Public Health, vol. 8, p. 250, 2020. https://doi.org/10.3389/fpubh.2020.00250.

[2] S. Wang, L. Jiang, J. Meng, Y. Xie, and H. Ding, "Training for smart manufacturing using a mobile robot-based production line," Frontiers of Mechanical Engineering, vol. 16, pp. 249–270, 2021. https://doi.org/10.1007/s11465-020-0625-z.

[3] A. Bhutoria, "Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model," Computers and Education: Artificial Intelligence, vol. 3, p. 100068, 2022. https://doi.org/10.1016/j.caeai.2022.100068.

[4] M. Liu and R. Su, "Research on the Path of Digital Transformation of Postgraduate Education in Chinese Universities under the Background of Digital Education Strategy," Intell Inf Manag, vol. 15, no. 5, pp. 339–349, 2023. https://doi.org/10.4236/iim.2023.155016.

[5] Y. Jiang and B. Li, "Exploration on the teaching reform measure for machine learning course system of artificial intelligence specialty," Sci Program, vol. 2021, pp. 1–9, 2021. https://doi.org/10.1155/2021/8971588.

[6] H. He, H. Yan, and W. Liu, "Intelligent teaching ability of contemporary college talents based on BP neural network and fuzzy mathematical model," Journal of Intelligent & Fuzzy Systems, vol. 39, no. 4, pp. 4913–4923, 2020. https://doi.org/10.3233/JIFS-179977.

[7] C. Liu, X. Li, and J. Zhang, "Construction of Students' Innovation Ability Portrait for Cultural Intelligence Computing," International Journal of Crowd Science, vol. 7, no. 2, pp. 77–86, 2023. https://doi.org/10.26599/IJCS.2023.9100001.

[8] M. Li, and Y. Su, "Evaluation of online teaching quality of basic education based on artificial intelligence," International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 16, pp. 147-161, 2020. https://www.learntechlib.org/p/217942/.

[9] L. Lee, and Y.-S. Liu, "Training effects and intelligent evaluated pattern of the holistic music educational approach for children with developmental delay," International Journal of Environmental Research and Public Health, vol. 18, no. 19, pp. 10064, 2021. https://doi.org/10.3390/ijerph181910064.

[10] L. P. Yang, and T. Xin, "Changing Educational Assessments in the Post - COVID - 19 Era: From Assessment of Learning (AoL) to Assessment as Learning (AaL)," Educational Measurement: Issues and Practice, vol. 41, no. 1, pp. 54-60, 2022. https://doi.org/10.1111/emip.12492.

[11] M. Wang, and Z. Lv, "Construction of personalized learning and knowledge system of chemistry specialty via the internet of things and clustering algorithm," The Journal of Supercomputing, vol. 78, no. 8, pp. 10997-11014, 2022. https://doi.org/10.1007/s11227-022-04315-8.

[12] H. Lee, and Y. Hwang, "Technology-enhanced education through VR-making and metaverse-linking to foster teacher readiness and sustainable learning," Sustainability, vol. 14, no. 8, pp. 4786, 2022. https://doi.org/10.3390/su14084786.

[13] A. R. Haniah, A. Aman, and R. Setiawan, "Integration of strengthening of character education and higher order thinking skills in history learning," Journal of Education and Learning (EduLearn), vol. 14, no. 2, pp. 183-190, 2020. https://doi.org/10.11591/edulearn.v14i2.15010.

[14] D. Denny and I. Iskandar, "The mastery of teacher emotional intelligence facing 21st century learning," International Journal of Education and Teaching Zone, vol. 1, no. 1, pp. 50–59, 2022. https://karya.brin.go.id/id/eprint/13167.

[15] Y. Baashar et al., "Evaluation of postgraduate academic performance using artificial intelligence models," Alexandria Engineering Journal, vol. 61, no. 12, pp. 9867–9878, 2022. https://doi.org/10.1016/j.aej.2022.03.021.

[16] S. Grabowska and S. Saniuk, "Assessment of the competitiveness and effectiveness of an open business model in the industry 4.0 environment," Journal of Open Innovation: Technology, Market, and Complexity, vol. 8, no. 1, p. 57, 2022. https://doi.org/10.3390/joitmc8010057.

[17] W. Wei and Y. Jin, "A novel Internet of Things-supported intelligent education management system implemented via collaboration of knowledge and data," Mathematical Biosciences and Engineering, vol. 20, no. 7, pp. 13457–13473, 2023. https://doi.org/10.3934/mbe.2023600.

[18] C. Malamateniou et al., "Artificial intelligence: guidance for clinical imaging and therapeutic radiography professionals, a summary by the Society of Radiographers AI working group," Radiography, vol. 27, no. 4, pp. 1192–1202, 2021. https://doi.org/10.1016/j.radi.2021.07.028.

[19] L. H. Son, and H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," Applied Intelligence, vol. 49, no. 1, pp. 172-187, 2019. https://doi.org/10.1007/s10489-018-1262-7.

[20] C.-Y. Huang, B.-Y. Cheng, S.-J. Lou, and C.-C. Chung, "Design and Effectiveness Evaluation of a Smart Greenhouse Virtual Reality Curriculum Based on STEAM Education," Sustainability, vol. 15, no. 10, pp. 7928, 2023. https://doi.org/10.3390/su15107928.

[21] H.-C. Tsai, A.-S. Lee, H.-N. Lee, C.-N. Chen, and Y.-C. Liu, "An application of the fuzzy Delphi method and fuzzy AHP on the discussion of training indicators for the regional competition, Taiwan national skills competition, in the trade of joinery," Sustainability, vol. 12, no. 10, pp. 4290, 2020. https://doi.org/10.3390/su12104290.

[22] L. H. Son, and H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," Applied Intelligence, vol. 49, no. 1, pp. 172-187, 2019. https://doi.org/10.1007/s10489-018-1262-7.

[23] S. S. Shreem, H. Turabieh, S. Al Azwari, and F. Baothman, "Enhanced binary genetic algorithm as a feature selection to predict student performance," Soft Computing, vol. 26, no. 4, pp. 1811-1823, 2022. https://doi.org/10.1007/s00500-021-06424-7.

[24] A. A. Yahya, "Swarm intelligence-based approach for educational data classification," Journal of King Saud University-Computer and Information Sciences, vol. 31, no. 1, pp. 35-51, 2019. https://doi.org/10.1016/j.jksuci.2017.08.002.

[25] Z. Zhu and L. Zhang, "Artificial Intelligence Empowers Postgraduate Education Ecologically Sustainable Development Model Construction," Sustainability, vol. 15, no. 7, p. 6157, 2023. https://doi.org/10.3390/su15076157.

[26] A. Thurzo, M. Strunga, R. Urban, J. Surovková, and K. I. Afrashtehfar, "Impact of artificial intelligence on dental education: a review and guide for curriculum update," Educ Sci (Basel), vol. 13, no. 2, p. 150, 2023. https://doi.org/10.3390/educsci13020150.

[27] X. Shan, J. Cao, and T. Xie, "The Development and Teaching of the Postgraduate Course 'Engineering System Modeling and Simulation' in Combination with Essentials Taken from Research Projects," Systems, vol. 11, no. 5, p. 225, 2023. https://doi.org/10.3390/systems11050225.

A Novel Fusion Deep Learning Approach for Retinal Disease Diagnosis Enhanced by Web Application Predictive Tool

Nani Gopal Barai[1], Subrata Banik[2], F M Javed Mehedi Shamrat[3]
Bangladesh Japan Information Technology Limited (BJIT Limited), Dhaka, Bangladesh[1, 2]
Department of Computer System and Technology, University of Malaya, Kuala Lumpur, Malaysia[3]

***Abstract*—Retinal disorders such as age-related macular degeneration and diabetic macular edema can lead to permanent blindness. Optical coherence tomography (OCT) enables professionals to observe cross-sections of the retina, which aids in diagnosis. Manually analyzing images is time-consuming, difficult, and prone to mistakes. In the dynamic and constantly evolving domain of artificial intelligence (AI) and medical imaging, our research represents a significant development in the field of retinal diagnostics. In this study, we introduced "RetiNet", an advanced hybrid model that is derived from the best features of ResNet50 and DenseNet121. To the model, we utilized an open-source retinal dataset that underwent a meticulous refinement process using a series of preprocessing techniques. The techniques involved Histogram Equalization for the purpose of achieving optimal contrast, Gaussian blur to mitigate noise, morphological operations to facilitate precise feature extraction, and Data Balancing to ensure impartial model training. These operations led to the attainment of a test accuracy of 98.50% by RetiNet, surpassing the performance standard set by existing models. A web application has been developed with the purpose of disease prediction, providing doctors with assistance in their diagnostic procedures. Through the development of RetiNet, our research not only transforms the accuracy of retinal diagnostics but also introduces an innovative combination of deep learning and application-oriented solutions. This innovation brings in a novel era characterized by improving reliability and efficiency in the field of medical diagnostics.**

Keywords—Retinal disease; RetiNet; hybrid model; learning; Web application; gaussian blur; histogram equalization

I. Introduction

This Retinal disorders commonly result from a combination of predispositions and environmental factors, mostly affecting the structures within the eyes, such as the membrane, lens, and nerve systems within eye. These conditions exert a significant impact on individuals' quality of life. Notably, approximately 7% of those aged 65 and above report experiencing a form of visual impairment [1]. The importance of timely and accurate diagnosis cannot be overstated in mitigating the severity of these conditions. The usage of inappropriate diagnostic approaches may exacerbate the issue. Diabetic Eye Disease (DED), comprising conditions such as diabetic retinopathy, glaucoma, and cataracts, is a significant cluster of visual impairment that impacts individuals with diabetes. Extended periods of diabetics can result in a decline in visual acuity and, in severe cases, substantial visual loss. As stated by the World Health Organization (WHO), among the global population of 2.2 billion people affected by visual impairments, it is estimated that approximately one billion cases may have been preventable with proper diagnostics and treatment [2].

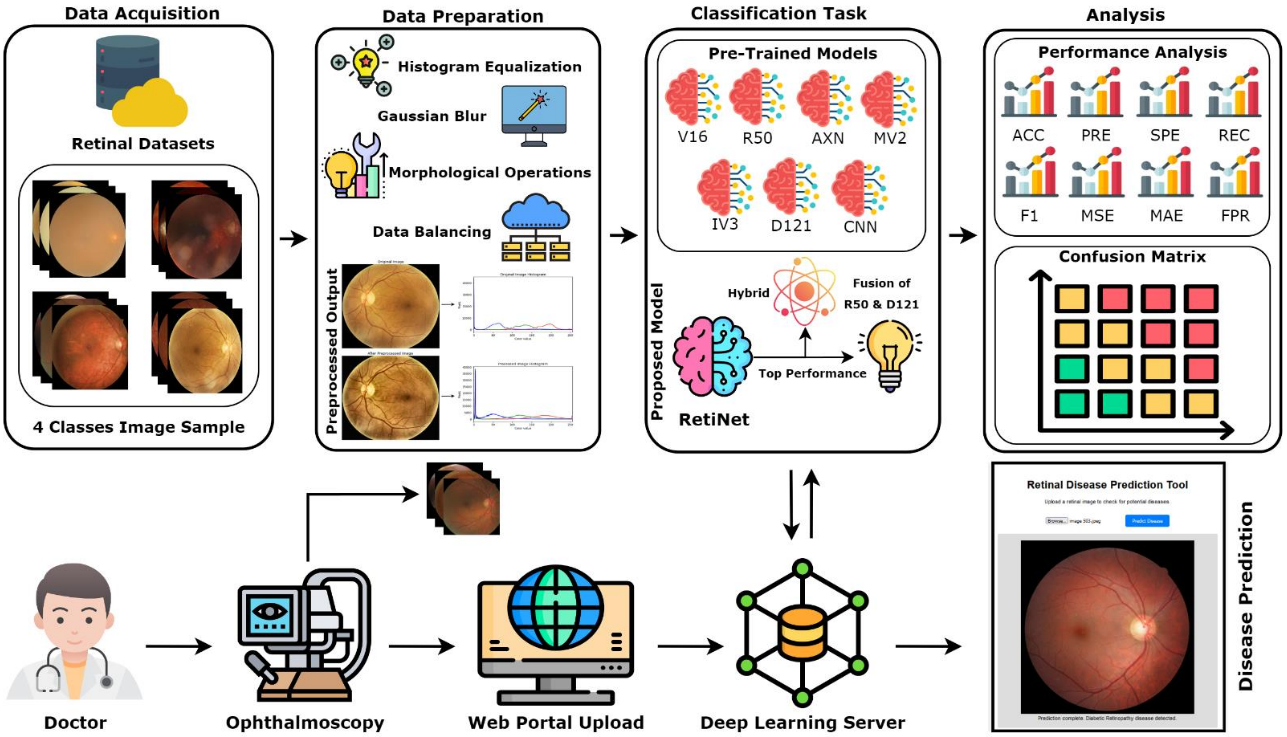
Artificial Intelligence has become an essential resource in assisting healthcare professionals in the early diagnosis of diseases [3, 4]. Currently, numerous AI-driven systems integrate medical test results with domain-specific knowledge to detect and categorize diseases. Furthermore, deep learning (DL) has been employed in various practical contexts, showcasing its promise. Specifically, researchers have utilized DL techniques to identify retinal disorders by analyzing retinal fundus images. Although DL approaches in the field of machine learning (ML) succeed at distinguishing between healthy and diseased retinal images, the challenge of classifying varied retinal disorders into multiple classes remains complex and unresolved.

Numerous studies have attempted to predict or characterize retinal health by analyzing eye images. For instance, the authors in [5] proposed an enhanced technique that employs a novel CenterNet model in conjunction with a DenseNet-100 feature extractor. The study aimed to identify and characterize lesions associated with diabetic retinopathy and macular edema. The approach demonstrated exceptional accuracies of 97.93% and 98.10% when assessed on the APTOS-2019 and IDRiD benchmark datasets. In a comparable way, the study conducted by [6] tested the performance of three distinct classification algorithms across various classes. These algorithms included the Convolutional Neural Network (CNN), Visual Geometry Group 16 (VGG16), and InceptionV3. The analysis and comparison of each approach were performed using the confusion matrix.

The prevalence of retinal disease in ophthalmology has been a significant concern due to its diverse manifestations and rising frequency. Traditional diagnostic approaches, although valuable, may prove ineffective due to inherited human fallibility, namely inaccurate assessment of the intricate details of retinal vision. In order to address challenges, we devised "RetiNet". RetiNet has been developed to surpass the existing limitations in the field of retinal image diagnosis by harnessing the potential of DL to address the associated challenges. The primary focus of RetiNet is the enhancement of diagnostic accuracy. Our study presents a comprehensive approach that presents cutting-edge AI techniques aiming to improve the standards of retinal disease detection. Fig. 1 represents the overall workflow of the proposed system. The primary findings of the study are as follows:

- This study introduced "RetiNet", a novel hybrid model generated from the fusion of ResNet50 and DenseNet121.
- A comprehensive set of preprocessing techniques were utilized, including Histogram Equalization, Gaussian Blur, Morphological Operations, and Data Balancing. The methods guaranteed optimal image quality and consistency, hence facilitating efficient model training and accurate diagnosis.
- The performance of RetiNet surpasses that of other popular models. The outcome emphasizes the efficiency of our hybrid methodology, demonstrating its potential for real-world implementation and reliability.
- This study incorporates a web application in a novel way, providing an efficient instrument for the diagnosis of retinal diseases.

The paper's structure is outlined in the following manner: Section II thoroughly examines previous research. Section III, which focuses on technique, addresses several issues such as data acquisition, data preparation, use of pre-trained models, proposed models, and their corresponding hyperparameters. Section IV provides an in-depth analysis of the results obtained from the research. Section V is on the development and implications of a functional web application. Section VI involves a discourse and comparative examination with prior research. Section VII eventually concludes the article by presenting essential findings and valuable perspectives.



Upload a retinal image to check for potential diseases.

Prediction complete. Diabetic Retinopathy disease detected.

Fig. 1. The overall architecture of the proposed hybrid model RetiNet.

II. Literature Review

Throughout several decades, the scientific community has diligently advocated for the advancement of automated diagnostic systems, with the ultimate goal of revolutionizing the field of medical diagnostics. Traditional expert systems, albeit groundbreaking, operated based on meticulously defined rules, occasionally encountering difficulties when confronted with complex classification challenges. Nevertheless, introducing machine and deep learning has brought forth a renewed vigor in this sector. Through the use of data training, these algorithms have made a substantial contribution to the improvement of research endeavors. Significantly, within the scope of medical research, machine learning distinguishes itself due to its remarkable versatility, demonstrating proficiency in detecting and classifying a multitude of disorders across various domains.

In a recent study by Arslan et al. [7], various CNN architectures were employed to evaluate a 2748 Retinal Fundus images dataset. This dataset comprised 1374 images from healthy individuals and 1374 images from diverse diseased groups. The CNN models underwent thorough evaluation using the 10-fold cross-validation methodology. Remarkably, the EfficientNet architecture demonstrated superior performance with an impressive accuracy and recall rate of 94.88% in measurements. Similarly, Malik et al. [8] developed a rapid diagnosis approach for eye diseases by utilizing a diverse range of machine learning models using Neural Networks. This array consisted of algorithms such as the Decision Tree, Random

Forest, and Naïve Bayes. The ICD-10 codes were used for specific disease diagnoses. Among the classifiers, the Random Forest model had superior performance, achieving an accuracy rate of 86.63%. The Neural Network model closely trailed after, achieving an accuracy rate of 85.98%. In a separate investigation centered on Optical Coherence Tomography, Metin and Karasulu [9] employed two CNN models, ResNet50 and MobileNetV2, as the foundation for their research. When the aforementioned models were employed to analyze data pertaining to several retinal diseases, ResNet50 exhibited an accuracy of 94% whilst MobileNetV2 dropped to 81%.

Sarki et al. [10] introduced an automated technique for the detection of Diabetic Eye Diseases (DEDs) by the analysis of retinal fundus images. The classifier chosen for their study was the CNN model, which was further optimized by fine-tuning using the RMSprop optimizer. The dataset, comprised of images from multiple sources, was specifically curated to facilitate multi-class classification. Remarkably, their classifier achieved an accuracy rate of 81.33% in retinal disease classification. Hussain et al. [11] conducted another study utilizing a dataset consisting of 251 spectral-domain optical coherence tomography (SD-OCT) images. Among the sample, a total of 192 cases were indicative of retinal diseases, while the remaining 59 cases were of healthy eyes. The researchers utilized the Random Forest classification technique to differentiate between data indicative of illness and normal conditions. To ensure the model's robustness, a 15-fold cross-validation process was utilized, yielding a remarkable classification accuracy of 96.89%.

Additionally, Almansour et al. [12] introduced a CNN architecture that drew inspiration from the VGG16 model, with a specific focus on glaucoma detection. The researchers acquired data from a total of seven distinct databases. The localization of the Region of Interest (ROI) was given particular emphasis, resulting in the allocation of a total of 2084 samples for classification subsequent to ROI analysis. To enhance the performance of the VGG16 model, two more layers were incorporated, ending with the Softmax activating function. Upon evaluating the combined data from all the datasets, the model demonstrated a noteworthy accuracy rate of 78%. In an independent study, Seker et al. [13] developed a CNN model using Keras framework for the purpose of glaucoma classification. The fundus images were first subjected to preprocessing using the Irfanview graphic schemes prior to being inputted into the classification pipeline. The suggested framework was robust, consisting of 49 layers and employing the Adam optimizer with binary cross-entropy loss function. This demonstrates a classification accuracy of 85%.

The limitations in the current retinal disease identification using conventional imaging techniques have been evident through various research. Numerous contemporary models encounter challenges pertaining to the accuracy rate of predictions and the complexity involved in augmenting retinal image datasets. In order to address these challenges and enhance the quality of retinal diagnostics, we introduced “RetiNet”. The primary objective with RetiNet is to utilize its hybrid architecture to achieve exceptional accuracy in the diagnosis and classification of retinal diseases. This attempt highlights our commitment to enhancing diagnostic skills and redefining the parameters of retinal imaging diagnostics.

III. Methodology

A. Data Acquisition

In the context of retinal disease diagnosis investigation, we employed an extensive dataset of retinal images collected from the renowned open-source platform Kaggle [14]. The dataset is crucial to our study as it contains many images corresponding to various eye conditions. This comprehensive collection allows for a balanced and intricate approach to our analysis. The dataset consists of 1074 images depicting Healthy Eyes, 1038 images of Cataract, 1098 images implying Diabetic Retinopathy, and 1007 images displaying Glaucoma. Such heterogeneity not only contributes to the robust classification of various eye diseases but also provides a dimension of authenticity to our investigation. For an understanding of the range of diversity present in the dataset, Fig. 2 may serve as a useful reference, as it displays sample images from each category. The utilization of this dataset has played a pivotal role in our research, facilitating the establishment of evaluations and findings based on empirical evidence from the real world.

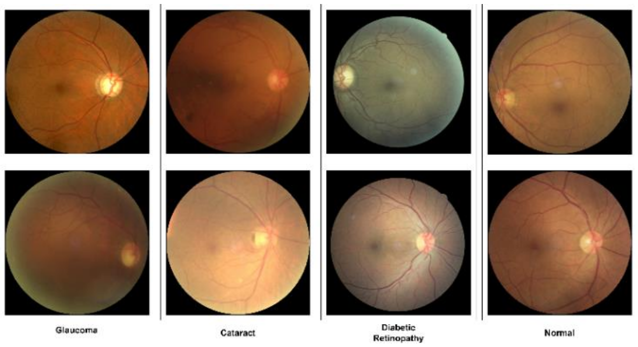


Fig. 2. Sample images of the four classes from the selected retinal datasets.

B. Data-preparation

1) Histogram equalization: The study utilized a specialized histogram equalization technique to manage variations in contrasts across color images effectively. The first step involved the conversion of each image from the RGB color system to the YCbCr color space. The application of histogram equalization was restricted solely to the Y channel, representing luminance, to mitigate any potential color distortion. The equalization procedure commenced by calculating the histogram for the Y channel, which signifies the distribution of pixel intensities. The resulting Cumulative Distribution function (CDF) was utilized to remap each pixel intensity, r in the Y channel according to the following formula:

$$s = \frac{(CDF(r) - CDF_{min})}{(1 - CDF_{min})}\ (L-1) \tag{1}$$

Here, L is the grayscale level, and CDFmin is the smallest non-zero value in the CDF.

This process facilitates the even dispersion of luminance intensities, consequently enhancing regions with low contrast. After equalization, the modified Y channel was reintegrated into the YCbCr image, which was subsequently transformed back to RGB. This approach resulted in a standardized contrast through the entire dataset, generating a consistent foundation for subsequent analysis. Algorithm 1 outlines the operational steps involved in the implementation of histogram equalization.

Algorithm 1: Histogram Equalization for Retinal Images

```
Procedure ColorHistEqu (Image $I$, GrayscaleLevels $L$)
    $I_{YCcCr} \leftarrow$ Convert $I$ to YCbCr color space
    Extract $Y$ channel as $I_Y$
    Initialize histogram array $H$ [0….$L$-1] to zeros
    Initialize CDF array $CDF$ [0….$L$-1] to zeros
    for each pixel $p$ in $I_Y$ do
        $H\ [intensity\ of\ P] \leftarrow H\ [intensity\ of\ P] + 1$
    end for
    $CDF\ [0] \leftarrow H\ [0]$
    for $i$ = 1 to $L$-1 do
        $CDF\ [i] \leftarrow CDF\ [i-1] + H[i]$
    end for
    $cdf_{min} \leftarrow$ minimum non-zero value of $CDF$
    for $i$ = 0 to $L-1$ do
        $normalized_{cdf}\ [i] \leftarrow \frac{(CDF[i]-cdf_{min})}{(1-cdf_{min})}\ (L-1)$
    end for
    for each pixel $p$ in $I_Y$ do
        $P_{newintensity} \leftarrow normalized_{cdf}[intensity\ of\ P]$
    end for
    Replace Y channel in $I_{YCbCr}$ with $I_Y$
    Convert $I_{YCbCr}$ back to RGB color spcae
    Save $I_{YCbCr}$ as 'Output.jpg'
    return $I_{YCbCr}$
end procedure
```

2) Gaussian blur: The Gaussian blur is a widely used convolutional technique where Gaussian Kernel G is convolved with an image I. In the domain of image processing, a technique akin to a weighted average of pixel values is employed, wherein the weights progressively diminish as the distance from the center pixel increases. The Gaussian function in two dimensions is mathematically defined in Eq. (2).

$$G(x,y) = \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

Here, the variables x and y represent spatial coordinates, whereas σ is the standard deviation that governs the extent of the Gaussian kernel's distribution. The kernel radius is commonly selected as μ =3 σ to account for more than 99% of the Gaussian distribution. To implement the Gaussian Blur on an image, a convolution operation is executed between the image and the Gaussian kernel.

$$(I * G)(u,v) = \sum_{x=-\mu}^{\mu}\sum_{y=-\mu}^{\mu} I(u-x, v-y)\cdot G(x,y) \tag{3}$$

The variable (u, v) represent the pixel coordinates in image I, whereas the variables (x, y) iterate over the dimensions of the Gaussian kernel. The outcome, denoted as $(I * G)(u, v)$, represents a blurred image.

The convolution process is employed to amplify the influence of the central pixels relative to the distant ones, resulting in a visible smoothing effect. As σ increases, the level of blurring becomes more pronounced, resulting in a greater impact on the surrounding pixels. Algorithm 2 is an illustration of the Gaussian blur technique.

Algorithm 2: Gaussian Blur for Image Smoothing

```
Procedure GaussianBlur(Image $I$, StandardDeviation $\sigma$)
    Compute Gaussian kernel radius $\mu \leftarrow 3\times\sigma$
    Initialize Gaussian Kernel G with size $(2\mu+1)\times(2\mu+1)$
    for $x = -\mu to \mu, y = -\mu to \mu$ do
        $G[x+\mu][y+\mu] \leftarrow \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}}$
    end for
    Normalize kernel $G$ such that its sum is 1
    for each pixel $(u, v)$ in $I$ do
        $I_{blurred}[u][v] \leftarrow \sum_{x=-\mu}^{\mu}\sum_{y=-\mu}^{\mu} I(u-x, v-y) \times$
        $G[x+y][y+\mu]$
    end for
    return $I_{blurred}$
end procedure
```

3) Morphological operations: The assessment of eye diseases using retinal images requires meticulous attention to image quality and clarity. Morphological methods, which are fundamentally non-linear in nature, have been integral to our efforts in enhancing these images based on their form. For instance, the dilation technique is used to amplify white regions of the foreground of images. The process of enlarging retinal images improves recognizable characteristics, particularly, blood vessels, to make them easier to identify when applied on images. The concept can be articulated as,

$$f \oplus s\ (x,y) = max_{(a,b)\in s}\{f(x-a, y-b)\} \tag{4}$$

In contrast, erosion functions as an inverse of dilation as it reduces the white regions in the images. This contraction is highly advantageous for the purpose of severing associated objects or eliminating minor noise components, as observed in the context.

$$f \ominus s\ (x,y) = min_{(a,b)\in s}\{f(x+a, y+b)\} \tag{5}$$

The improvement of images can be achieved by implementing the opening operation, which involves a sequential process of erosion followed by dilation. This technique effectively eliminates minor protrusions or objects, which is a necessary tool for reducing noise and artifacts. This operation can be represented as,

$$f \circ s = (f \ominus s) \oplus s \tag{6}$$

The final component of the morphological approaches is the closure procedure. Starting with dilating, followed by degrading an image, it adeptly closes tiny holes or breaches in the foreground, required to restore the discontinuities in blood vessels. The aforementioned procedure can be depicted as,

$$f \bullet s = (f \oplus s) \ominus s \tag{7}$$

Following the series of preprocessing techniques, such as histogram Equalization, Gaussian Blur, and Morphological Operations, the retinal images underwent processing to achieve optimal clarity and enhancement of features. The images generated are prepared for training and subsequently fed into our neural network model, as illustrated in Fig. 3. The histogram of both images (original and preprocessed) shows the changes after applying all the preprocessing techniques (Histogram Equalization, Gaussian blur, and morphological operations).

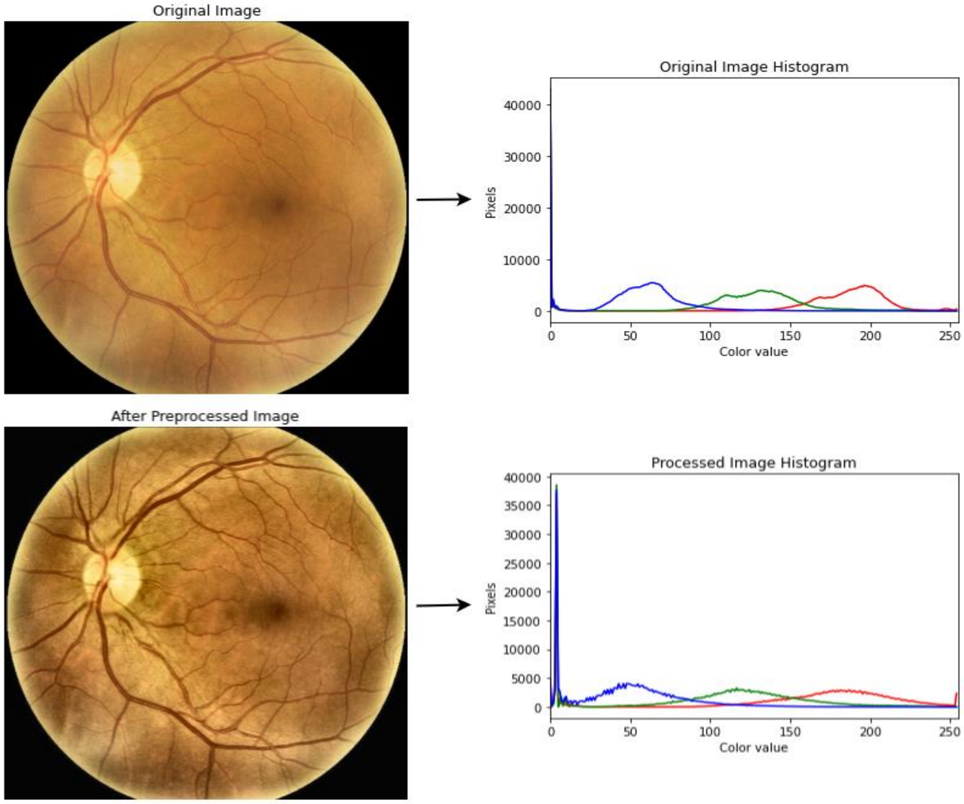


Fig. 3. After all preprocessed steps histogram showing the changes of image enhancement ratio.

4) Data balancing: In the domain of machine learning and data analysis, the presence of an unbalanced dataset has the potential to introduce bias into the model predictions, particularly when one class is disproportionately represented in comparison to the others. The disparity between the classes was detected in our study by analyzing retinal images. The data under-sampling method was employed to resolve this issue. Strategic undersampling involves reducing the number of instances in the overrepresented or majority classes in order to align with the size of the minority class rather than artificially increasing the number of instances in the underrepresented class. This approach equalized 1000 images for each class covering Normal, Diabetic Retinopathy, Cataract, and Glaucoma. While implementing this strategy decreased the overall amount of data for the majority classes, it played a vital role in achieving a balanced distribution of data. A balanced dataset reduces the potential for biased predictions while simultaneously improving the model's capacity for generalization. Algorithm 3 illustrates the general under-sampling technique.

Algorithm 3: Data Undersampling for Class Balance

```
Procedure Undersample (Dataset $D$)
    Determine the size of the smallest class, $minSize$
    for each class $c$ in $D$ do
        $data_c \leftarrow$ Data instances of class $c$
        if size of $data_c > minSize$ then
            Randomly select $minSize$ instances from $data_c$
            to form $newdata_c$
        else
            $newdata_c$
        end if
    end for
    Merge all $newdata_c$ to form the balanced dataset $D'$
    return $D'$
end procedure
```

After establishing a balanced dataset, we further split our data into training, test, and validation subsets to assist the modeling stage. The division adopted a ratio of 70:20:10. The training set consists of 70% of the images from each class, serving as the fundamental data for model learning. The test set comprised 20% of the data and was used to evaluate the performance of the training model on previously unseen data.

The remaining 10% of the data was allocated as the validation set, which serves as an essential component for repeated modification and fine-tuning of the model during the training process. Table I presents a comprehensive analysis detailing the distribution of images across each set and class. The table presented visually demonstrates the rigorous distribution method adopted to achieve effective model training and validation.

TABLE I. FINAL DATA DISTRIBUTION FOR EACH CLASS AND THE ALLOCATION OF DATA INTO TRAINING, TESTING, AND VALIDATION SETS

Class	Original	Balance
Normal	1074	1000
Diabetic Retinopathy	1098	1000
Cataract	1038	1000
Glucoma	1007	1000
Total	4217	4000
Train	2800 Images (70%)	
Test	800 Images (20%)	
Validation	400 Images (10%)	

C. Model Evaluation and Classification

The primary objective of our proposed approach is to automatically identify eye disorder while delivering an improved level of classification accuracy. To address the challenges of classifying retinal conditions from retinal datasets, we have developed a novel and robust hybrid CNN model RetiNet. This model is characterized by its distinctiveness, reliability and resilience. In the pursuit of identifying the most effective transfer learning approach for the classification task at hand, we tested seven recognized pre-trained models: VGG16 [15], ResNet50 [16], AlexNet [17], MobileNetV2 [18], InceptionV3 [19], DenseNet121 [20] and a CNN [21]. Table II provides an overview of the important characteristics and fundamental elements of the chosen deep CNN models. The subsequent sections provide a comprehensive evaluation of the architecture and performance of RetiNet, as well as the primary focus of the study.

TABLE II. FEATURES AND ATTRIBUTES OF EVALUATED DEEP LEARNING MODELS

Model	Input Shape	Custom Input Shape	Parameters	Size (MB)
VGG16	224×224	224×224	138,35,7544	528
ResNet50	224×224	224×224	25,636,712	98
AlexNet	227×227	224×224	62,378,344	233
MobileNetV2	224×224	224×224	35,38,984	14
InceptionV3	229×229	224×224	23,851,784	92
DenseNet121	224×224	224×224	80,62,504	33
CNN	224×224	224×224	78,81,365	39
RetiNet	224×224	224×224	35,100,000	136

1) Proposed Models (RetiNet): In the context of eye disease classification from retinal images, a sensitive field of study, the deliberate and evidence-driven decision to combine the capabilities of ResNet50 and DenseNet121. ResNet50 is renowned for its innovative skip or “residual connections” which efficiently counteracts the vanishing gradient issues that often arise in deep neural networks. This ensures that, regardless of the depth of the network, gradients flow smoothly, hence facilitating rapid learning. This ResNet has consistently exhibited its efficacy in many image classification challenges, successfully detecting both broad structures, such as the overarching form of blood vessels, as well as the intricate details, such as subtle deviations that may indicate potential disorders. In contrast, DenseNet121 exhibits an exceptionally thick architecture, where each layer establishes intimate interconnections with all other layers, facilitating a seamless transmission of information.

Since minor details in retinal images might function as early indicators of disorders such as diabetic retinopathy or glaucoma, the dense linkage encourages feature reuse. Additionally, it functions as a built-in regularization mechanism, safeguarding against overfitting, a critical consideration when dealing with constrained datasets. While architectures such as VGG and Inception possess their own merits, our hybrid model is particularly well-suited for the complex demands of retinal image classification due to the combination of the depth and skip connections of ResNet50, along with the extensive feature extraction capabilities of DenseNet-121. The proposed RetiNet model is outlined as follows:

When presented with an input image I, both the ResNet50 and DenseNet121 architectures execute a series of convolutional operations to extract feature maps.

$$F_{resnet}(I) = ResNet50\ (I) \tag{10}$$

$$F_{densenet}(I) = DenseNet121\ (I) \tag{11}$$

where, F_{resnet} and $F_{densenet}$ are the feature maps from ResNet50 and DenseNet121, respectively.

Furthermore, we apply global average pooling (GAP) to these feature maps to get a fixed-size feature vector. For a given feature map F, the GAP operation can be expressed as,

$$GAP\ (F) = \frac{1}{W \times \mathrm{H}} \sum_{i=1}^{W} \sum_{j=1}^{H} F(i,j) \tag{12}$$

where, W and H are the width and height of the feature maps, respectively.

However, the feature vectors obtained from the GAP operation on both networks are concatenated.

$$C = Concatenate\ (GAP(F_{resnet}), GAP(F_{densenet})) \tag{13}$$

The concatenated feature vector C is passed through a dense layer with a ReLU activation function,

$$D = \sigma(W_d \times \mathrm{C} + b_d) \tag{14}$$

where, σ is the ReLU activation function, W_d represents the weight of the dense layer and $b_d$is the bias.

Finally, the feature vector from the dense layer is passed through another dense layer with softmax activation to classify the image into the four classes of eye disease.

$$O = Softmax(W_o \times \mathrm{D} + b_o) \tag{15}$$

where, W_o represents the weights of the output layer and b_o is the bias.

The output O will be a vector with four values, each representing the probability of the image belonging to the corresponding class (Normal, Diabetic Retinopathy, Cataract, and Glaucoma). Algorithm 4 demonstrates the procedural steps of the proposed RetiNet model. Furthermore, Fig. 4 illustrates the architecture of the hybrid model.

Algorithm 4: RetiNet: Hybrid Model for Eye Disease Classification

```
Procedure RetiNet (Image I)
    $I_{norm} \leftarrow$ Normalize $I$
    $F_R \leftarrow$ ResNet50_Extract ($I_{norm}$)
    $F_D \leftarrow$ DenseNet121_Extract ($I_{norm}$)
    $G_R \leftarrow$ GAP ($F_R$)
    $G_D \leftarrow$ GAP ($F_D$)
    $C \leftarrow Concat\ (G_R, G_D)$
    $D \leftarrow Dense\_ReLU\ (C)$
    $O \leftarrow Softmax\ (D)$
    return $O$
end procedure
```

D. Hyperparameters Optimization

Hyperparameters are significant variables that might have an influence on the training dynamics and overall performance of the model [22], [23]. The variables include the number of epochs, batch size, image dimensions, optimizer options, activation functions, learning rate, decay rate, dropout rate, and regularization parameters. During the experiment, we repeatedly modified several parameters, including batch size, learning rate, and regularization variables, resulting in improvements in the model's accuracy and efficiency. RetiNet was subjected to benchmarking against several prominent architectures, namely VGG16, ResNet50, AlexNet, MobileNetV2, InceptionV3, DenseNet121, and a customized CNN. The training process for each model consisted of 300 epochs, during which various optimizers were employed to facilitate a comprehensive evaluation. To fine-tune our model and to determine the optimal hyperparameter configuration, we apply the keras-tune tool, followed by an extensive grid search technique. Table III presents the final set of hyperparameters post-tuning.

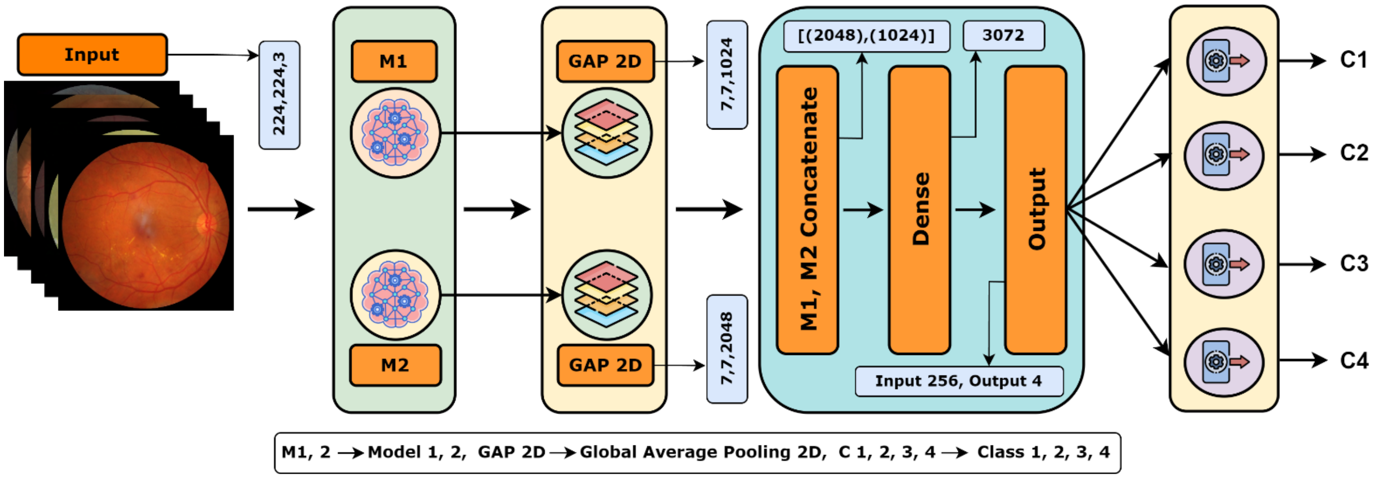


Fig. 4. The architecture of the proposed model RetiNet.

TABLE III. OPTIMIZED HYPERPARAMETERS FOR MODEL TRAINING

Model	No. of Epochs	Batch Size	Image Size	Optimizers	Activation Function	Learning Rate	Decay Rate	Dropout Rate	Regularizer
VGG16	300	64	224×224	Adam	Softmax	0.000001	1e-3	0.5	5e-4
ResNet50	300	64	224×224	SGD	ReLU	0.0001	1e-4	-	1e-4
AlexNet	300	64	224×224	Adagrad	ReLU	0.00001	1e-2	0.2	5e-4
MobileNetV2	300	64	224×224	SGD	Softmax	0.1	1e-4	0.5	1e-5
InceptionV3	300	64	224×224	Adam	Sigmoid	0.001	1e-3	-	1e-4
DenseNet121	300	64	224×224	Adam	ReLU	0.0000001	1e-2	-	1e-4
CNN	300	64	224×224	Adam	ReLU	0.001	1e-2	0.5	1e-4
RetiNet	300	128	224×224	RMSProp	Sigmoid	0.001	1e-6	0.9	1e-4

IV. ANALYSIS OF EXPERIMENTAL OUTCOMES

A. Environmental Setup and Tools

We implemented all the deep learning models using Keras (version 2.10.0) and TensorFlow (version 2.0) frameworks using Python 3.7. For data visualization, we deployed the Seaborn and Matplotlib packages. The evaluation was carried out on a device powered by an AMD Ryzen 7 CPU clocked at 3.90 GHz, 32 GB RAM, and an AMD Radeon RX 580 series GPU, functioning on a Windows 10 operating system.

B. Assessment Metrics

To comprehensively evaluate the performance of our model in the classification of retinal images, we adopted a set of statistical metrics that involves Accuracy, Specificity, Recall, Precision, False Positive Rate (FPR), F1-score, Mean Squared Error (MSE), and Mean Absolute Error (MAE).

Accuracy: Represents the proportion of correct predictions made by the model over the total predictions.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{16}$$

Here, TP = True Positives, TN= True Negatives, FP= False Positives and FN = False Negatives.

Precision: Indicates the fraction of relevant instances among the instances that the model predicted as positive. It provides insight onto the correctness of positive predictions.

$$Precision = \frac{TP}{TP+FP} \tag{17}$$

Recall: Signifies the proportion of real positive occurrences that the model managed to predict accurately, demonstrating its ability to detect positive cases.

$$Recall = \frac{TP}{TP+FN} \tag{18}$$

Specificity: Measures the actual negative rate, indicating the model's effectiveness in accurately recognizing the negative class amongst all the classes.

$$Specificty\ = \frac{TN}{TN+FP} \tag{19}$$

FPR: Proportion of negative instances that are incorrectly classified as positive.

$$FPR = \frac{FP}{FP+TN} \tag{20}$$

F1-Score: Provides a balance between Precision and Recall by combining them into a single measure. This metric also shows how well a model can be used to identify both positive and negative data.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{21}$$

MSE: Determines the average squared variance between the predicted outcomes and the actual values, delivering a sense of prediction error size.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y^{\wedge}{}_i)^2 \tag{22}$$

Here y_i is the actual value, $y^{\wedge}{}_i$ is the predicted value, and n is the number of observations.

MAE: Provides the mean of the absolute variance between predictions and actual observations, exhibiting the model's accuracy.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y^{\wedge}{}_i| \tag{23}$$

C. Study Outcomes

In this study, eight algorithms were utilized to classify retinal image data. These algorithms comprised seven transfer learning models and the novel RetiNet model. These models aim to assist in identifying eye conditions by accurately detecting abnormalities in retinal images. In order to conduct an extensive evaluation, each model underwent training for 300 epochs, with the results being recorded at each iteration. The performance measures of each model can be calculated by utilizing Eq. (16) to Eq. (23). This thorough methodology provides a profound understanding of the functioning of each model when exposed to retinal data.

This study systematically evaluated the performance of eight distinct models on their ability to classify retinal images with high consistency. Fig. 5 illustrates the performance of all the models. Among the employed transfer learning models, VGG16 exhibited commendable performance, achieving an accuracy of 91.58% and a precision of 92.1%. These results serve as evidence of its proficiency in accurately classifying positive instances. The model, ResNet50, achieved an accuracy of 86.36% and a specificity of 86.7%, indicating its ability to classify negative samples effectively. The AlexNet model, with its accuracy and recall rates of 92.76% and 92.9%, respectively, demonstrated its adeptness in identifying actual positive rates.

The MobileNetV2 and InceptioV3 models exhibited distinct performance characteristics, achieving accuracies of 93.31% and 93.5%, respectively. Both models displayed impressive F1 scores, representing the harmonic mean of accuracy and recall, indicating a well-balanced performance across both metrics. InceptionV3 demonstrated a balanced detection capability, as evidenced by the achievement of an F1-score of 93.35%. When comparing the performance of DenseNet121 and CNN, it is observed that both models achieved satisfactory results, with accuracies of 90.86% and 88.34%, respectively. These findings suggest that there may be potential for further enhancement in the model's performance. The accuracy rate of the improved model, RetiNet is 98.50%. Another notable statistical measure that evaluates the average squared variations between predicts and actual observations, known as the Mean Squared Error (MSE), was found to be exceptionally low for RetiNet at 0.015. This indicates that the predictions made by RetiNet are highly accurate and closely aligned with the actual results.

Nevertheless, the focal point of the assessment was the cutting-edge RetiNet model. The system exhibits a remarkable level of accuracy of 98.50% and F1-score of 98.65%. Additionally, it demonstrates a precision of 98.7% and an impressively low FPR of 1.7%. This accomplishment showcases the proficiency of RetiNet in accurately identifying and distinguishing retinal abnormalities with high precision.

The combination of all these measurements offers an exhaustive overview of the capabilities of each model. The detailed analysis depicted in Fig. 5 not only highlights the potential of models such as RetiNet but also paves the path for future advancements in the field of retinal image diagnostics.

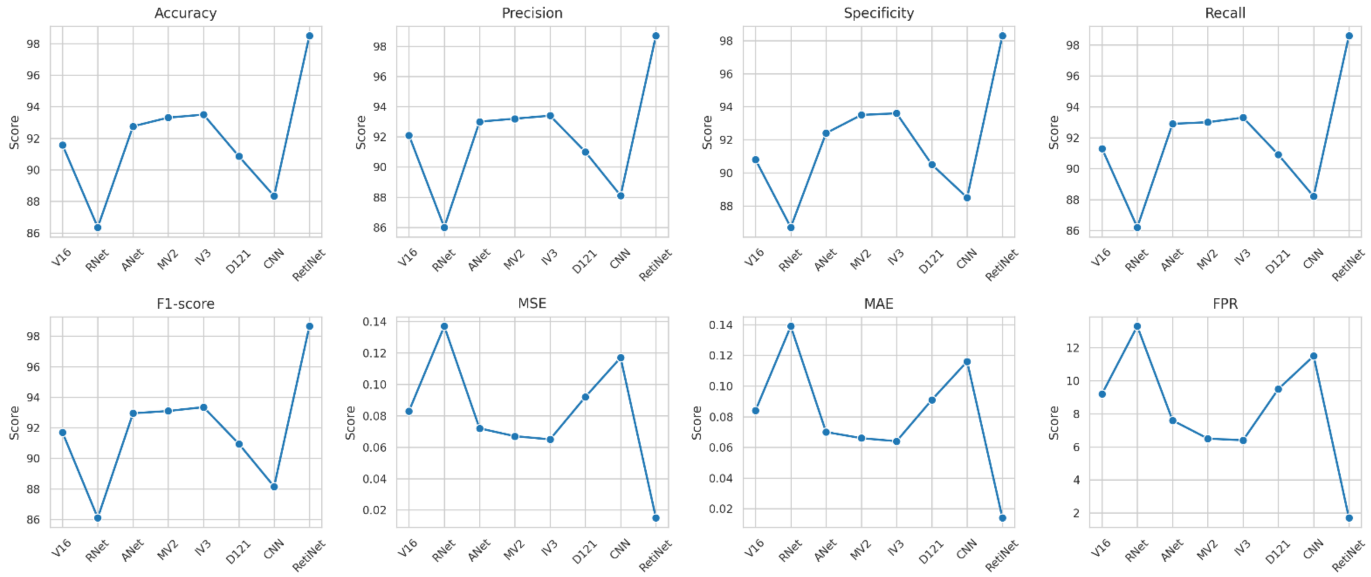


Fig. 5. Comparative performance matrics of retinal image classification models. Here ‘V16’ indicates ‘VGG16’, ‘RNet’ indicates ‘ResNet50’, ‘Anet’ indicates ‘AlexNet’, ‘MV2’ indicates ‘MobileNetV2’, ‘IV3’ indicates ‘InceptionV3’, ‘D121’ indicates ‘DenseNet121’.

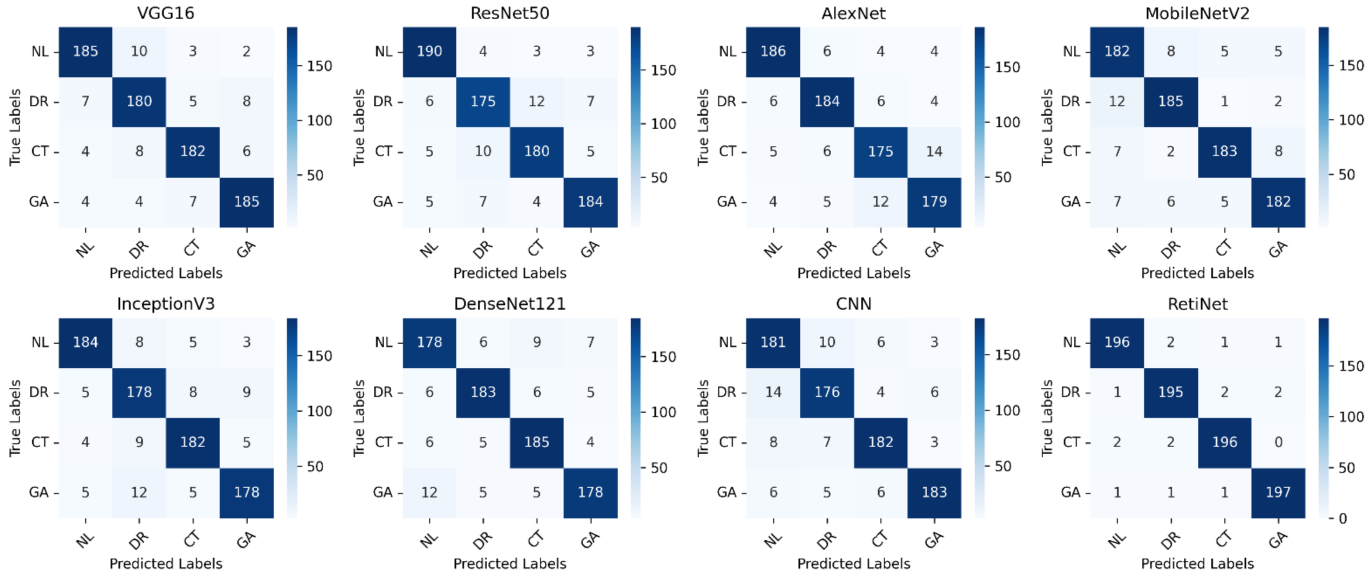


Fig. 6. Confusion matrix of all classification models. Here ‘NL’ indicates ‘Normal’ class, ‘DR’ indicates ‘Diabetic Retinopathy’ class, ‘CT’ indicates ‘Cataract’ class, GA indicates ‘Glaucoma’ class.

During the testing phase, the model's performance is evaluated using confusion matrices and observing significant patterns of classification and misclassification illustrated in Fig. 6. Out of the 200 Normal data, the VGG16 model accurately classified 185 data while misclassifying 15 data. More specifically, 10 of the data was misclassified as Diabetic Retinopathy. In the Diabetic Retinopathy class, 180 images were correctly identified, whereas five images were incorrectly categorized as Cataract. On the other hand, 182 images depicting Cataract were accurately identified, whereas eight images were erroneously interpreted as Diabetic Retinopathy. Lastly, 185 images were accurately classified in the Glaucoma class, yet seven were mislabeled as Cataracts.

The ResNet50 model accurately classified 190 images as normal and misclassified 10 images, of which four were incorrectly labeled as Diabetic Retinopathy. In the class focused on Diabetic Retinopathy, 175 images were accurately recognized, whereas 12 were mislabeled as Cataract. Among the images in the Cataract class, 180 were found to be correct, while five were classified as Glaucoma. Finally, for the Glaucoma class, 184 accurate classifications were made, and four images were misclassified as Cataract.

AlexNet properly classified 186 images as Normal, while six images as misdiagnosed as Diabetic Retinopathy. In the case of Diabetic Retinopathy, 184 images were correctly classified, except six were misclassified as Cataract. From the

Cataract class, a total of 175 images were correctly identified, and 14 were wrongly identified as Glaucoma. A total of 179 images from the Glaucoma class were classified, and 12 were misclassified as Cataract. The model MobileNetV2 demonstrated proficiency in correctly classifying 182 images as Normal but misclassified eight as Diabetic Retinopathy. The condition known as Diabetic Retinopathy found 185 accurate classifications, but 12 images were mistakenly classified as Normal. The model correctly classified 183 Cataract images but labeled 8 as Glaucoma. Similarly, 182 Glaucoma images could be correctly predicted but seven images were misinterpreted as Normal.

In the case of the model InceptionV3, 184 Normal images were correctly classified and eight were classified as Diabetic Retinopathy. The model identified 178 images of Diabetic Retinopathy but incorrectly identified eight images as Cataract. Within the Cataract class, 182 images were diagnosed, but 5 were misdiagnosed as Glaucoma. From the Glaucoma class, 178 instances were identified, while 12 instances were misidentified as Diabetic Retinopathy. The DenseNet121 properly identified 178 Normal images while incorrectly identifying six as Diabetic Retinopathy. The model accurately classified 183 Diabetic Retinopathy images yet six images wrongly fell into the Cataract category. Within the Cataract class, a total of 185 images were correctly recognized, while four images were erroneously classified as Glaucoma. In the Glaucoma class, there were 178 accurate predictions and 12 instances where the classification was incorrectly assigned as Normal.

In the CNN model, 200 Normal images were tested, of which 181 were correctly classified, while 10 were inaccurately classified as Diabetic Retinopathy. The model properly recognized 176 Diabetic Retinopathy images while incorrectly reporting 14 as the Normal group. For cataract, 182 were accurately recognized, with three misinterpreted for Glaucoma. In the Glaucoma category, 183 were correctly classified, with five misclassified as Diabetic Retinopathy.

Finally, the RetiNet model, the standout performer, successfully identified an outstanding 196 out of 200 Normal images, with only two misclassifications into the Diabetic Retinopathy group. It also properly detected 195 Diabetic Retinopathy images, with misinterpretation of two images into the Cataract class. In the cataract group, 196 interpretations were correctly made, with only two misclassifications into the Glaucoma class. The Glaucoma class witnessed 197 valid classifications with just a single misclassification into the Cataract class. The performance of all the eight models employed in the study is presented in Table IV.

TABLE IV. PERFORMANCE SCORES OF ALL EMPLOYED MODELS FOR RETINAL DISEASE DIAGNOSIS

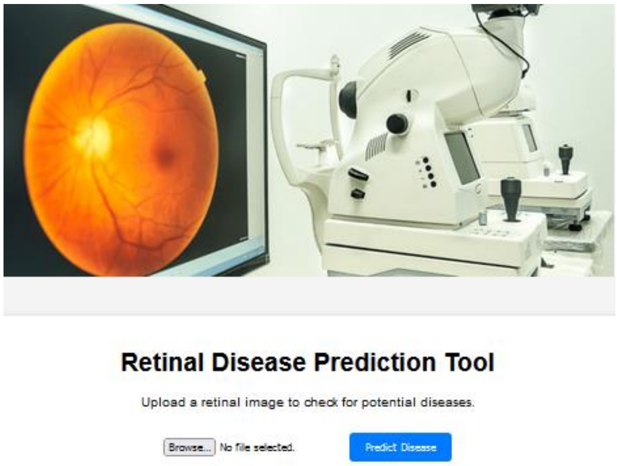
Model	VGG16	ResNet50	AlexNet	MobileNetV2	InceptionV3	DenseNet121	CNN	RetiNet
Accuracy	91.58%	86.36%	92.76%	93.31%	93.50%	90.86%	88.34%	98.50%
Precision	92.10%	86.00%	93.00%	93.20%	93.40%	91.00%	88.10%	98.70%
Specificity	90.80%	86.70%	92.40%	93.50%	93.60%	90.50%	88.50%	98.30%
Recall	91.30%	86.20%	92.90%	93.00%	93.30%	90.90%	88.20%	98.60%
F1-score	91.70%	86.10%	92.95%	93.10%	93.35%	90.95%	88.15%	98.65%
MSE	0.083	0.137	0.072	0.067	0.065	0.092	0.117	0.015
MAE	0.084	0.139	0.070	0.066	0.064	0.091	0.116	0.014
FPR	9.2%	13.3%	7.6%	6.5%	6.4%	9.5%	11.5%	1.7%

V. WEB INTERFACE

A precisely designed digital interface has been built exclusively for medical professionals, with a focus on facilitating their use of medical imaging for diagnosing different retinal diseases. The implementation of this modern interface plays a crucial role in facilitating healthcare professionals, particularly doctors, and experts, in effectively and accurately identifying the medical issues impacting their patients. Fig. 7 displays the general interface of the web application.

Doctors initiate the procedure by entering medical images of the patient's retina into the system. The images are instantly uploaded to the server and merged into the proposed RetiNet diagnostic model. Concurrently, the diagnostic form that goes with it is filled out with essential patient data and sent to the server again.

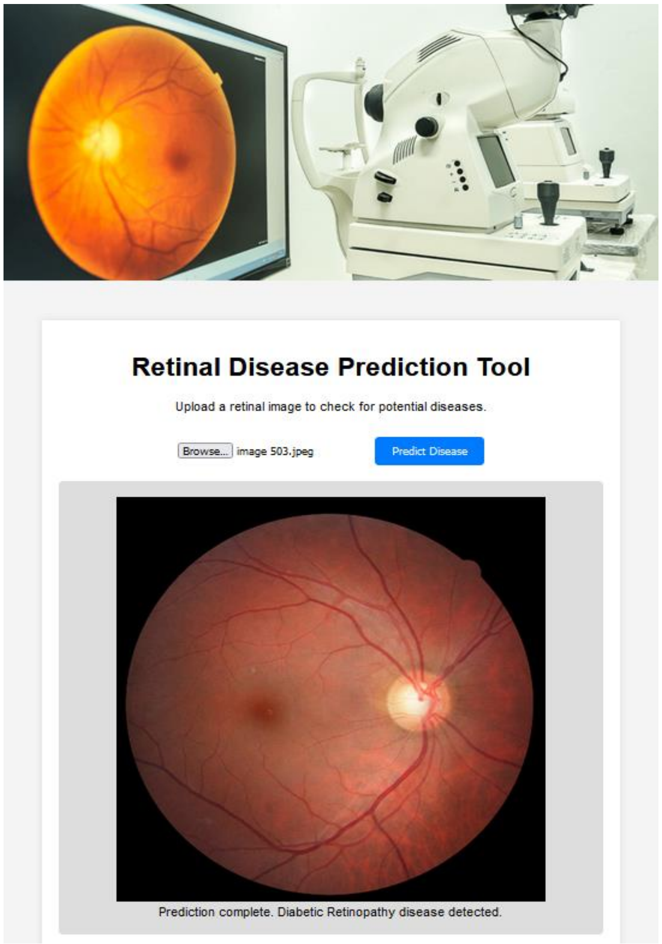
During this phase, the server conducts an essential verification process evaluating the quality of the retinal images that have been submitted to ensure they satisfy the requisite criteria for precise analysis. After conducting this verification process the server utilizes proposed RetiNet technology to examine the images and generate an advanced AI supported estimation of the stage of the disease. This process entails employing of an advanced algorithmic interpretation of the retinal images to determine the condition's progression and severity.



Retinal Disease Prediction Tool

Upload a retinal image to check for potential diseases.

Fig. 7. Web application interface.


Retinal Disease Prediction Tool

Upload a retinal image to check for potential diseases.

Prediction complete. Diabetic Retinopathy disease detected.

Fig. 8. Disease diagnosis interface.

After the analysis is accomplished, the server aggregates the findings, which include the AI-generated estimate of disease's stage. The doctor is then instantly informed of these findings. Fig. 8 shows the prediction of AI technology help doctors diagnose patients more accurately with more expertise.

VI. Comparison of Existing Studies

This study presents the "RetiNet" model, an advanced CNN model specifically developed for accurately classifying retinal images. One of the primary objectives is to solve the significant challenge of accurately diagnosing retinal disease within the imaging datasets. Prior to training, the dataset was subjected to thorough preprocessing techniques. These techniques involve Histogram Equalization for optimum image contrast, Gaussian Blur for noise reduction, and Morphological Operations for enhanced feature extraction and Data Balancing to offset class imbalances. During the testing phase, RetiNet demonstrates a remarkable accuracy of 98.50%, surpassing other existing models and effectively enabling disease detection. Moreover, A web-based tool has been created to assist medical professionals in detecting and diagnosing retinal disorders. This initiative set the benchmark for retinal image-based diagnosis and provides medical practitioners with a robust diagnostic tool. Table V provides a detailed analysis of the performance of RetiNet in comparison to other prominent models.

TABLE V. Proposed Model Comparison of Existing Studies

Authors	Methods	Accuracy
Arslan et al. [7]	EfficientNet	94.88%
Malik et al. [8]	Random Forest	86.63%
Metin and Karasulu [9]	ResNet50	94.00%
Sarki et al. [10]	CNN	81.33%
Hussain et al. [11]	Random Forest	96.89%
Almansour et al. [12]	Fine-tuned VGG16	78.00%
Seker et al. [13]	Keras-based CNN	85.00%
Barai et al.	Proposed Model (RetiNet)	98.50%

VII. Conclusion

This study discusses the significant advantages of employing retinal images for accurately identifying retinal diseases through utilizing our cutting-edge "RetiNet" model. The study subjects each image to a comprehensive preprocessing procedure that includes Histogram Equalization, Gaussian Blur, Morphological Operations, and Data Balancing. We have obtained the highest degree of classification for images that are well-suited for the purpose of extracting features efficiently. A web application has been designed and developed to assist medical professionals in identifying retinal diseases. Although RetiNet's ability did detect some misclassification, its overriding performance spotlighted the transformational potential of CNN models in retinal imaging. Further studies will prioritize the improvement of these methods in order to increase preciseness, with the ultimate goal of transforming the diagnosis of retinal diseases worldwide and significantly influencing ophthalmic healthcare.

References

[1] Pelletier, A. L., Rojas-Roldan, L., & Coffin, J. (2016). Vision loss in older adults. American family physician, 94(3), 219-226.

[2] World Health Organization (2019), https://www.who.int/newsroom/newsletters, (Last Access: 02.04.2023).

[3] Shamrat, F. J. M., Azam, S., Karim, A., Ahmed, K., Bui, F. M., & De Boer, F. (2023). High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. Computers in Biology and Medicine, 155, 106646.

[4] Shamrat, F. J. M., Akter, S., Azam, S., Karim, A., Ghosh, P., Tasnim, Z., ... & Ahmed, K. (2023). AlzheimerNet: An effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images. IEEE Access, 11, 16376-16395.

[5] Nazir, T., Nawaz, M., Rashid, J., Mahum, R., Masood, M., Mehmood, A., Ali, F., Kim, J., Kwon, H. & Hussain, A. Detection of Diabetic Eye Disease from Retinal Images Using a Deep Learning Based CenterNet Model. Sensors. 21 (2021), https://www.mdpi.com/1424-8220/21/16/5283.

[6] Smaida, M. & Serhii, Y. Comparative Study of Image Classification Algorithms for Eyes Diseases Diagnostic. International Journal Of Innovative Science And Research Technology. 4 (2019).

[7] ARSLAN, G., & Erdaş, Ç. B. (2023). Detection of Cataract, Diabetic Retinopathy and Glaucoma Eye Diseases with Deep Learning Approach. Intelligent Methods In Engineering Sciences, 2(2), 42-47.

[8] Malik, S., Kanwal, N., Asghar, M. N., Sadiq, M. A. A., Karamat, I., & Fleury, M. (2019). Data driven approach for eye disease classification with machine learning. Applied Sciences, 9(14), 2789.

[9] METİN, B., & KARASULU, B. (2022). Derin Öğrenme Modellerini Kullanarak İnsan Retinasının Optik Koherans Tomografi Görüntülerinden Hastalık Tespiti. Veri Bilimi, 5(2), 9-19.

[10] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2021). Convolutional neural network for multi-class classification of diabetic

eye disease. EAI Endorsed Transactions on Scalable Information Systems, 9(4).

[11] Hussain, M. A., Bhuiyan, A., D. Luu, C., Theodore Smith, R., H. Guymer, R., Ishikawa, H., ... & Ramamohanarao, K. (2018). Classification of healthy and diseased retina using SD-OCT imaging and Random Forest algorithm. PloS one, 13(6), e0198281.

[12] Almansour, A., Alawad, M., Aljouie, A., Almatar, H., Qureshi, W., Alabdulkader, B., ... & Almazroa, A. (2022). Peripapillary atrophy classification using CNN deep learning for glaucoma screening. Plos one, 17(10), e0275446.

[13] Seker, M. E., Koyluoglu, Y. O., Celebi, A. R. C., & Bayram, B. (2022). Effects of Open-Source Image Preprocessing on Glaucoma and Glaucoma Suspect Fundus Image Differentiation with CNN.

[14] Kaggle. Avialble Online: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification (Accessed on 10 August 2023).

[15] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[17] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[18] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[19] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[20] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[21] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[22] Akter, S., Shamrat, F. J. M., Chakraborty, S., Karim, A., & Azam, S. (2021). COVID-19 detection using deep learning algorithm on chest X-ray images. Biology, 10(11), 1174.

[23] Sutradhar, A., Al Rafi, M., Ghosh, P., Shamrat, F. J. M., Moniruzzaman, M., Ahmed, K., ... & Moni, M. A. (2023). An Intelligent Thyroid Diagnosis System Utilising Multiple Ensemble and Explainable Algorithms with Medical Supported Attributes. IEEE Transactions on Artificial Intelligence.

Enhancing Production System Performance: Failure Detection and Availability Improvement with Deep Learning and Genetic Algorithm

Artika Farhana[1], Shaista Sabeer[2], Ayasha Siddiqua[3], Afsana Anjum[4]
Lecturer, Department of Computer Science, Jazan University, Saudi Arabia[1]
Lecturer, Department of Computer Science & Information Technology, Jazan University, Saudi Arabia[2, 3]
Lecturer, Department of Computer Science & Information Technology, Jazan University Jazan, KSA[4]

Abstract*—A crucial component of industrial operations is the detection of production system failures, which aims to spot any problems before they get worse. By applying cutting-edge methods like deep learning and genetic algorithms, failure detection accuracy may be improved, allowing for preemptive actions to reduce downtime and maximize system availability. These methods improve reactivity to possible errors and solve dynamic issues, which enhances the overall efficiency and reliability of production systems. This study offers a novel method for improving the availability and failure detection of production systems using deep learning techniques and genetic algorithms in a data-driven strategy. The goal of the project is to provide a complete framework for efficient failure detection that incorporates deep learning models, particularly Convolutional Neural Network (CNN) Autoencoder. Furthermore, system configurations are optimized through the use of genetic algorithms, improving overall availability. The suggested model is able to identify complex patterns and connections in the data by being trained on a variety of datasets that contain information about equipment failure. The incorporation of genetic algorithm guarantees flexibility and resilience in system setups, hence augmenting total availability. The study presents a proactive and flexible approach to the dynamic issues encountered in industrial environments, providing a notable breakthrough in failure detection and availability improvement. The proposed model is implemented in Python software. It achieves an astounding 99.32% accuracy rate, which is 3.58% higher than that of current techniques like CNN-LSTM (Long Short-Term Memory), Bi-LSTM (Bi-directional Long Short-Term Memory), and CNN-RNN (Recurrent Neural Network). The data-driven approach's high accuracy highlights its efficacy in forecasting and avoiding problems, which minimizes downtime and maximizes production efficiency.

***Keywords*—Autoencoder; availability enhancement; convolutional neural network; failure detection; genetic algorithm**

I. Introduction

Improving the performance of production systems on an ongoing basis is a primary goal in the manufacturing and industrial domains [1]. It is a result of the necessity to maximize productivity, reduce downtime, and guarantee the reliable provision of high-quality goods and services. Production systems must adapt to the competitive environment by moving beyond conventional paradigms and embracing novel approaches in order to remain robust in the face of shifting consumer expectations and technology breakthroughs [2]. This study explores the crucial area of production system improvement in an effort to push the envelope by presenting cutting-edge tactics that go beyond accepted constraints [3]. It begins with the planning and design of the product, moves through the complex manufacturing procedures, and concludes with the distribution and pleasure of the consumer [4].

The intrinsic difficulties that production systems encounter, such as the requirement to strike a balance between competing demands like cost-effectiveness, quality control, and on-time delivery, necessitate improvement [5]. Reactive methods of system optimization, which deal with problems as they arise, frequently lead to inefficient use of resources and more downtime [6]. The emphasis is shifting towards proactive and predictive techniques that foresee problems before they become serious ones as industries seek for leaner, more flexible production processes. In order to improve production system performance, this research will critically assess current approaches and offer novel solutions that make use of state-of-the-art technology [7]. It attempts to close the knowledge gap between theoretical developments and real-world application by offering sectors looking to stay competitive in a time of quickening technology change and shifting consumer expectations practical insights [8].

The early detection and mitigation of possible system failures in production systems represent key problems in the context of industrial operations and critical infrastructure [9]. Conventional failure detection techniques can result in expensive downtime and inefficiencies since they are reactive in nature and frequently rely on rule-based systems or simple algorithms. Recognizing this, the current research offers a deep learning-based data-driven approach for failure detection, which represents a revolutionary paradigm change [10]. Using deep learning to uncover complex patterns from large datasets, this methodology deviates from traditional methods and improves the system’s capacity to anticipate and proactively handle failure situations. [11].

Deep learning has the potential to improve failure detection accuracy and provide the necessary scalability for the complicated production situations of today [12]. Improving availability is a complex task that goes beyond maintaining system uptime. It entails striking a tactical balance between the effectiveness of maintenance, system dependability, and flexibility in response to shifting operating circumstances [13]. In order to identify optimal solutions within the large parameter space of a production system, genetic algorithms, as an optimization tool, replicate the process of evolution and provide a possible path towards overcoming these obstacles.

Conventional methods of accomplishing these goals frequently depend on reactive tactics, which deal with problems only after they occur, resulting in downtime and inefficient use of resources [14]. Genetic algorithms are incorporated into the framework as a complement to the deep learning paradigm in order to optimize system availability. Natural selection and evolution serve as the inspiration for genetic algorithms, which offer a potent way to discover and develop the best possible configurations for the production system. These algorithms constantly alter parameters, searching for configurations that minimize failure risk and maximize total system availability through recurrent processes of crossover, mutation, and selection. This study’s combination of genetic algorithms and deep learning offers a clever and comprehensive method for managing production systems. Through the integration of deep learning’s predictive powers with genetic algorithms’ optimization skills, the suggested framework seeks to create a proactive system that can both detect possible malfunctions and modify the system configuration to improve overall availability. The study is at the vanguard of efforts to transform the monitoring, maintenance, and optimization of production systems due to the synergy between innovative data-driven technologies.

Thekey contributions of the article is,

- The CNN Autoencoder based on deep learning is incorporated for failure detection. By automatically extracting pertinent features from the input data, the CNN Autoencoder improves the model's ability to identify complex patterns that may indicate probable problems, offering a reliable and data-driven method of failure detection.
- The Genetic Algorithm is employed to improve availability by optimizing system configurations in response to changing operating circumstances. This genetic algorithm-based method guarantees flexibility and resilience, tackling the intricacies of industrial production processes and leading to increased availability overall.
- The integration of CNN Autoencoder and Genetic Algorithm, results in a comprehensive approach for availability improvement and failure detection. By using cutting-edge data-driven methodologies to systematically handle both failure detection and system optimization elements, this all-encompassing strategy improves the overall resilience and efficiency of production systems.

There are five primary sections of the paper: In Section II, relevant works in the subject of production system failure detection are reviewed and current techniques are summarized. The issue statement is defined in Section III, along with the shortcomings and difficulties of the existing methods. The suggested technique, which combines genetic algorithms and deep learning to provide an advanced data-driven solution, is described in Section IV. The model's output is shown in Section V, along with a discussion of the model's performance and accuracy in relation to other techniques. The report is finally concluded in Section VI, which summarizes the main conclusions and offers directions for further research into the field of production system failure detection.

II. Related Works

In light of climate change and sustainability issues, this study discusses the pressing need for smart energy production and highlights the shortcomings of conventional first-principle model-based strategies in an environment of growing system scale and uncertainty [15]. The article provides a thorough analysis and emphasizes the ways in which Data-Driven Control (DDC) and Machine Learning (ML) approaches are used to the tracking, regulating, optimizing, and fault-detection of power production facilities. It provides a thorough analysis of the ways in which these cutting-edge techniques help to resolve ambiguities and improve the efficiency of both traditional thermal power generation and renewable energy sources. In regards to visibility, maneuverability, adaptability, economic viability, and safety, the article lists the benefits of ML and DDC. It is crucial to remember that the study does not go into great detail about the particular difficulties or restrictions related to the application of ML and DDC procedures to power production structures, leaving space for additional investigations to investigate the possible downsides and improve these strategies for real-world use. The inherent drawback of data-driven modeling becomes apparent when confronted with the constraints of both online and offline data, including issues such as data incompleteness stemming from loss, uncertainty, and bias.

In the framework of ring spinning technological advances, the study presents a unique method to preventative care with an emphasis on forecasting and health monitoring [16]. It suggests utilizing predictive analytics to create a data-driven preventive service system built on a regularized Deep Neural network (DNN). To keep an eye on crucial parts, the method makes use of a system of sensors that is built into the frames of spinning machines, each of which has many spindles. A GA is developed for multi-sensor assessments and prediction, demonstrating its efficiency in leveraging bigger amounts of data with comparatively small training data sets. With the use of a neural sensor network, the framework provides condition-based evaluation for every component in order to anticipate anomalies, disruptions, and failures in real time. The study does not, however, go into great detail on the shortcomings or difficulties of the suggested model, providing opportunity for more study to investigate such limits and improve the execution for more widespread industrial uses.

In order to forecast failure behavior in the industrial environment, the study investigates the use of sophisticated data analytics, more especially a data-driven Failure Mode and Effective Analysis (FMEA) approach [17]. Utilizing operational and historical data from investment in industrial items' usage stage, the technique applies DL models to improve maintenance scheduling visibility and decision assistance. A real-world scenario in the aviation industry is used to verify the structure, and the results show an astounding 95% accuracy in defect prediction. By incorporating these findings into a data-driven FMEA framework, the dependency on employee knowledge and skill is lessened, and variability in risk and failure probability predictions is eliminated. Nevertheless, the study does not go into great detail about any drawbacks or difficulties with the suggested technique, providing opportunity for more research into real-world implementation problems and wider application across other industrial settings.

This study explores the intricacies of semiconductor production, which is a multistep process that uses a variety of equipment's and sub processes to create miniaturized electronic circuits [18]. With the goal of improving the process of producing semiconductors, the study uses data mining, SPC, and data-driven decision-making frameworks to analyze production process information in great detail. The goal is to use the newest technologies powered by data to increase productivity and reliability. The study highlights the potential for major process enhancements and offers a thorough analysis of current procedures; however, it does not go into great length on the drawbacks or difficulties that come with using methods that use data in the production of semiconductors. Some issues are not completely addressed, such as possible biases, execution difficulties, and the adaptation of these approaches to varied production contexts. To evaluate the applicability and constraints of these methods based on data in a wider range of semiconductor production environments, more investigation is necessary.

This study tackles the critical problem of estimating RUL for equipment predictions, highlighting the importance of precise forecasts in reducing maintenance expenses and improving system dependability[19]. This study presents a new data-driven methodology called Convolutional Neural Network- Long-Short Term Memory- Particle Swarm Optimization (CNN–LSTM–PSO) hybrid DNN, which integrates conventional neural networks, LSTM networks, and CNN. In order to increase the accuracy of Remaining Useful Life (RUL) forecasting, this hybrid model seeks to identify spatial relations from time series with multiple variables data and retain nonlinear properties. To improve the efficiency of the network, the research uses PSO to optimize the network's hyper parameters. The suggested CNN–LSTM–PSO model is noteworthy for its ability to provide multi-step-ahead recommendations. Utilizing a NASA-provided lithium-ion battery dataset, the validation test shows that the CNN–LSTM–PSO model outperforms other cutting-edge ML and Deep Learning (DL) techniques when evaluating a variety of efficiency metrics. To allow for more research into the suggested model's relevance across various datasets and commercial situations, the report nevertheless fails to go into great detail about its drawbacks or possible drawbacks. Further investigation might examine the model's applicability to other machinery kinds and operating environments.

This paper explores the topic of smart production, with a particular emphasis on utilizing Artificial Intelligence (AI), ML, and sophisticated data analysis to optimize semiconductors production processes [20]. It emphasizes how Industrial Internet of Things (IIOT) sensors are being used in production more and more, which means effective data management is required. To solve issues in semiconductor production, the study suggests a dynamic method that combines algorithms for neural networks with evolutionary programming. In particular, the study presents a novel feature selection technique employing neural networks and evolutionary algorithms to improve the production process efficiency. Although the research offers a detailed analysis and innovative approaches, it does not fully address any potential drawbacks or difficulties related to the suggested dynamic algorithms and smart feature selection. Additional investigation is necessary due to practical issues, the algorithms' adaptation to a variety of industrial contexts, and possible limitations. Subsequent investigations have to evaluate the resilience and constraints of these suggested remedies in diverse semiconductor production environments and take into account practical implementation difficulties.

In order to anticipate the initial yields in semiconductor production, the research presents a combined structure that focuses on finding the best companion combinations for the Final Testing (FT) procedure [21]. In order to create an efficient prediction of yield model, the process entails converting categorical information into multivariate vectors using the entity anchoring technique and assessing several ML methods. A Genetic Algorithm (GA) incorporated in the yield prediction framework is used to find the optimal accessory combinations, optimizing for the greatest initial yield estimation, after the best-fit ML model has been identified. By combining the strengths of ML and GA, a smart prediction method is created that stabilizes the Overall Equipment Effectiveness (OEE) for the FT process and reduces the negative effects of improper accessory pairings on yield rates. But the report does not go into great detail about any drawbacks or difficulties with the proposed design, thus there is need for more research to examine concerns of adaptability, versatility, and practical application in various semiconductor production settings. To evaluate the combined framework's resilience and generality in various production contexts, more research is necessary.

In addition to highlighting reliability's critical role in modern production facilities, the article also addresses reliability's influence on systems lives, expenses for upkeep, and repair charges [22]. Although a number of reliability modelling approaches have been investigated, including Fault Trees, Petri Nets, and Markov Chains, the process of developing dependability models is still demanding of labor and dependent on expertise. The report suggests using data from contemporary manufacturing facilities for automation or assist in the creation of dependability models as a solution to this problem. With an emphasis on information-driven reliability evaluation for cyber-physical machines, the suggested methodology seeks to capitalize on the abundance of data produced in sophisticated production lines. A case study is included to test and improve the suggested data-driven strategy from a practical standpoint. Nevertheless, the study does not go into great detail about the possible drawbacks or difficulties with the suggested structure, providing opportunity for further investigation into real-world application problems, potential exaggerations in the data, and the applicability of the method in various manufacturing contexts. To evaluate the security and sustainability of the data-driven resilience evaluation in actual manufacturing environments, more research is necessary.

In order to increase system accessibility and lower life cycle costs, the study tackles the crucial job of forecasting the RUL of systems [23]. Using numerous sensor time series indications, a Deep Long-Short Term Memory (DLSTM) network-based technique is introduced in this suggested solution. Through the usage of its DL framework, the DLSTM model is intended to fuse the aforementioned signals in order to provide precise predictions for RUL by revealing latent long-term relationships. The DLSTM's attributes and network layout are effectively tuned for accurate and reliable predictions in this article using an adaptive moment assessment approach and a grid-based searching strategy. Utilizing two turbofan engine datasets, the DLSTM model's efficacy is verified, showing favorable outcomes when contrasted to other neural network algorithms and the latest methods documented in the available literature. Though there is room for additional studies to examine real-world difficulties in implementation, possible prejudices in the data, and the framework's flexibility to various system forms and operational circumstances, the article does not go into great detail about potential drawbacks or difficulties connected with the recommended DLSTM model. To evaluate the DLSTM model's adaptability and generalization in actual manufacturing environments, more research is necessary.

The essential subject of forecasting RUL in diverse engineering and manufacturing scenarios is the focus of the literature review. The use of sophisticated deep learning and machine learning methods, such LSTM networks, is a recurring topic in the development of precise and effective RUL forecasts. These models make use of many sensor time series signals, which makes it possible to identify complex patterns and hidden connections in the data. The importance of RUL prediction in raising system availability, cutting life cycle costs, and improving maintenance plans is emphasized in the studies. To ensure optimal performance, model parameters are often tuned using grid search techniques and adaptive algorithms. Experimental validation on various datasets, such as turbofan engines, consistently shows these models to perform as robustly and competitively against other neural network designs and state-of-the-art methodologies. All of these studies have one thing in common, though: they refrain from going into great detail about the difficulties, biases, and real-world implementation problems that come with using these sophisticated predictive models in industrial settings. This leaves space for future research to address these important issues.

III. Problem Statement

The current problem in industrial environments is the inefficiencies and disruptions brought about by unanticipated breakdowns in production systems, which result in more downtime, less efficient use of resources, and weakened system availability overall. These problems are not fully addressed by traditional reactive techniques to failure detection and system optimization. In order to address this issue, this research offers a novel solution: a deep learning-based, data-driven method for failure detection and availability augmentation that makes use of genetic algorithms. The challenge at hand is creating a comprehensive approach that uses genetic algorithms to dynamically optimize system configurations for increased availability and deep learning techniques to proactively identify failure antecedents in production systems. The goal is to revolutionize the way that production system management is now done by offering a proactive, adaptable framework that not only foresees problems before they happen but also continuously improves the system to maximize availability and overall performance [24]. Because of its ability to improve pattern recognition in industrial data—especially in identifying minor symptoms of equipment failure—the suggested CNN Autoencoder-GA approach has been chosen. The utilization of Convolutional Neural Network (CNN) Autoencoder enables efficient recognition of intricate patterns, and the incorporation of Genetic Algorithms (GA) guarantees the adaptability and durability of system configurations in ever-changing industrial settings. By training on a variety of datasets, the data-driven approach improves the model's resilience and adaptability. On the other hand, the shortcomings of current approaches, like insufficient pattern recognition, static configurations, and insufficient forecasting abilities, make them less appropriate for handling the dynamic issues associated with effective failure detection and availability enhancement in operational systems.

IV. Proposed Cnn Autoencoder – Genetic Algorithm Framework

In this paper, a unique data-driven technique for improving production system availability and failure detection is presented. The suggested system uses state-of-the-art techniques, such as deep learning and evolutionary algorithms, and integrates CNN Autoencoder for accurate failure detection. System configurations are optimized via genetic algorithms, increasing overall availability. The model, which has been trained on a variety of datasets, recognizes intricate patterns, and evolutionary methods guarantee system configuration flexibility. The model, which is implemented in Python, emphasizes how effective the data-driven approach is at issue prediction and prevention, downtime minimization, and production efficiency maximization. It is depicted in Fig. 1.

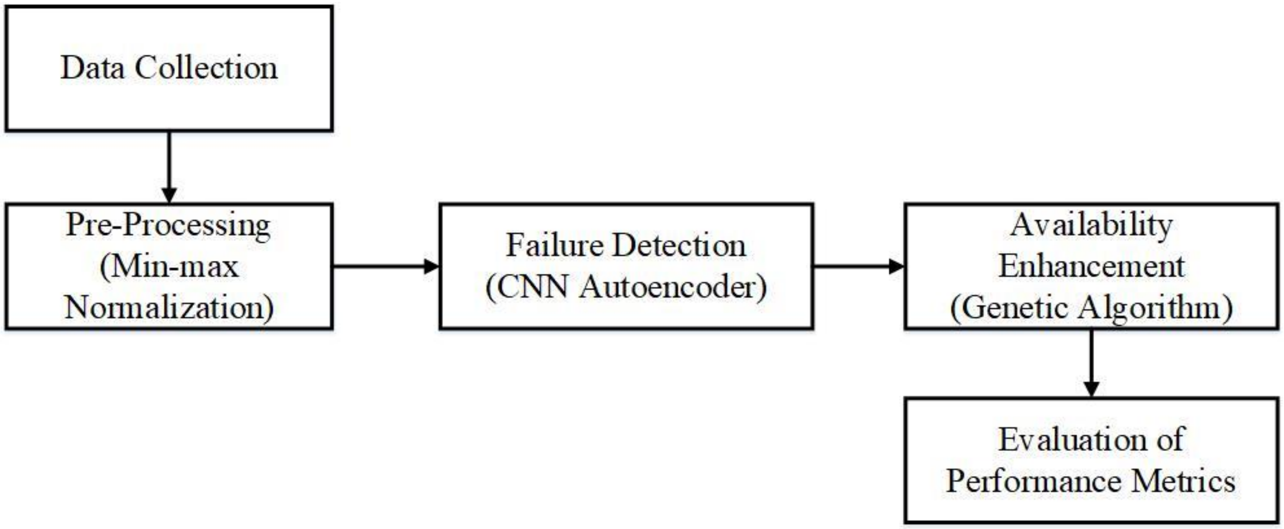

Data Collection
Pre-Processing (Min-max Normalization)
Failure Detection (CNN Autoencoder)
Availability Enhancement (Genetic Algorithm)
Evaluation of Performance Metrics

Fig. 1. Proposed methodology.

A. Data Collection

Kaggle, a well-known website for data science and machine learning contests, provided datasets for the equipment failure prediction. The meticulously selected datasets, which are publicly accessible on Kaggle, are an invaluable tool for practitioners and academics who are trying to create predictive models that may detect any malfunctions in machinery. The datasets available in Kaggle's repository cover a wide range of sectors and machinery kinds, making it possible to investigate various failure patterns and advance the creation of reliable prediction algorithms. These datasets are a great source of data for developing and accessing machine learning models, and they often contain elements like timestamps, operating parameters, and sensor readings. Researchers may test their models against pre-existing datasets by utilizing Kaggle's equipment failure prediction dataset platform. This approach promotes cooperation and creativity in the domain of predictive maintenance and reliability engineering [25].

B. Preprocessing using Min-Max Normalization

Preprocessing is essential when attempting to improve production system performance using a deep learning-based, data-driven strategy for genetic algorithm-based failure diagnosis and availability augmentation. In particular, applying Min-Max Normalization sticks out as an essential step in bringing the input data into compliance. This method makes sure that every variable contributes equally to the learning process by scaling the feature values within a given range, usually between 0 and 1. Normalizing the input data makes the deep learning model less susceptible to changes in the magnitude of various characteristics and more resilient, which supports steady and efficient training. By reducing the effect of different scales among input characteristics, Min-Max Normalization enhances the overall dependability of the model. This is important when it comes to failure detection and availability enhancement in intricate production systems. Moreover, the use of Min-Max Normalization is consistent with the overall objective of maximizing the accuracy and speed of convergence of the deep learning model. A more effective learning process is made possible by the standardized data distribution, which also helps to avoid particular traits from predominating during the training phase and thereby distorting the model's predictions. When it comes to availability enhancement and failure detection, wherein subtle patterns may signal approaching problems, and when accuracy is critical, the preprocessing step of Min-Max Normalization guarantees that the deep learning model is capable of identifying pertinent patterns and trends, which in turn enhances the model's ability to improve production system performance as shown in Eq. (1).

$$X_{Normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

C. Deep Learning-based CNN Autoencoder for Failure Detection

In the field of industrial system failure detection, this study presents a new method based on Deep Learning using CNN Autoencoder architecture. The main goal is to create a complicated model that can learn nuanced patterns from large, complex information in order to proactively anticipate probable problems. A change from conventional techniques is marked by the use of CNN Autoencoders, which use neural networks' capacity to automatically extract pertinent characteristics from input data. The CNN design is very useful for jobs that need spatial connections, which makes it suitable for examining images, sensor data, and other multidimensional data sources that are frequently found in industrial settings.

An encoding phase that compresses input data into a latent representation and a decoding phase that reconstructs the input from this compressed representation are the fundamental workings of the CNN Autoencoder. The model gains the ability to reflect typical operating patterns throughout training, which makes it sensitive to variations that might signal impending problems. Deep learning guarantees a data-driven and adaptable approach to failure detection, able to identify intricate patterns that may defy conventional rule-based or heuristic techniques. Using this cutting-edge method, the research hopes to enhance the creation of reliable and effective failure detection systems that may improve the performance and dependability of industrial production systems.

The encoder, which is composed with a number of layers of convolution, one or more fully connected layers, and a pooling procedure make up the CNN autoencoder. The decoder is proportionately made up of convolutional and upsampling layers after either two or three fully linked layers. In contrast to the fully connected autoencoder, the CNN autoencoder functions on a series of R includes vectors rather than just one. To keep things simple, let's look at an encoder and a decoder like the ones in Fig. 2 that have one convolutional layer and one fully linked layer. The convolutional component of the encoder works with an input matrix Y ∈ R B×R, wherein R is the sequence's frames count. The layer yields a B × R × C tensor H, the components of which are computed as shown in Eq. (2).

$$\text{K(a,b,c)} = \text{g} \; (\textstyle\sum_{m=0}^{h_{a-1}} \sum_{r=0}^{h_{b-1}} Y(a+m, b+n) h_c(m,n) \qquad (2)$$

where, Y(a, b) is the component at row a and column b of matrix Y, C is the total amount of kernels, g(•) is the non-linear activating operation, and $h_c$is a two-dimensional kernel of size $h_a \times h_b$. Keep in mind that because of zero-padding, K's initial measurements are the identical as Y's. Furthermore, although the operation carried out in the aforementioned equation is really a cross-correlation, the ML group generally refers to it as "convolution to operate.

Convolutional layers are frequently succeeded by a pooling process that lowers the input's complexity. The max-pooling procedure, that determines the highest value across a q × q windows, was utilized in (3):

$$\widetilde{K}\text{(a,b,c)} = \max \; \{\text{K(a',b',c')} : \text{a'} \in [\text{ a . s, a . s+q-1}], \text{b'} \in \; [\text{b . s, b . s+q-1}] \qquad (3)$$

where the step is located. It utilized q = 2 and s = 2 in the present study. A fully linked layer makes up the encoder's last layer.

The decoder is proportionately made up of a fully connected layer as the first layer, a layer that performs convolution, and an upsampling layer that replicates the input matrix's rows and columns using the identical factor 2, in this case that was utilized throughout pooling. In order to achieve two identical dimensions for the final output matrix as X, the system's final layer consists of one kernel and a convolutional layer with a linear activating function.

Applying additional convolutional and fully connected layers to the encoder and subsequently to the decoder would enhance the network's overall depth. In the trials, the real architecture was ascertained employing a validation set.

D. Employing Genetic Algorithm for Availability Enhancement

Towards strengthening the availability of production systems, the study deliberately employs Genetic Algorithms (GAs) as a sophisticated optimization method. Fundamentally, the main issue being addressed is the necessity of a methodical and flexible strategy to improve system availability by means of dynamic configuration optimization. Inspired by the ideas of evolution and natural selection, genetic algorithms are a powerful tool for negotiating the large solution space present in production systems' complexity. Potential system configurations are encoded into a population of people inside this optimization framework, each of which represents a distinct solution to the optimization issue. These solutions' fitness is carefully assessed using predetermined goals that are especially designed to improve system availability. This comprehensive strategy guarantees that the evolutionary algorithm converges to configurations that optimize the production system's overall availability while simultaneously reducing the likelihood of failures.

People in the population go through selection, crossover, and mutation processes as part of an iterative process known as the optimization journey. Individuals are chosen for development based on their fitness, and genetic material is transferred through crossover to produce progeny. Stochastic variations brought about by mutation encourage variety among the population. Motivated by availability-related goals, the fitness evaluation serves as a compass, pointing the algorithm in the direction of configurations that are reliable in reducing the likelihood of failure. The Genetic Algorithm becomes a powerful tool for converging across multiple generations to achieve configurations that dynamically adapt to the changing operational landscape, which eventually results in a production system optimized for increased resilience and availability.

Every individual in the population or possible solution is represented by a chromosome, frequently in binary form. Y_i, where*i* is the person's index, can be thought of as an individual solution in Eq. (4).

$$Y_i = (y_{i1}, y_{i2}, y_{i3}, \ldots, y_{in}) \qquad (4)$$

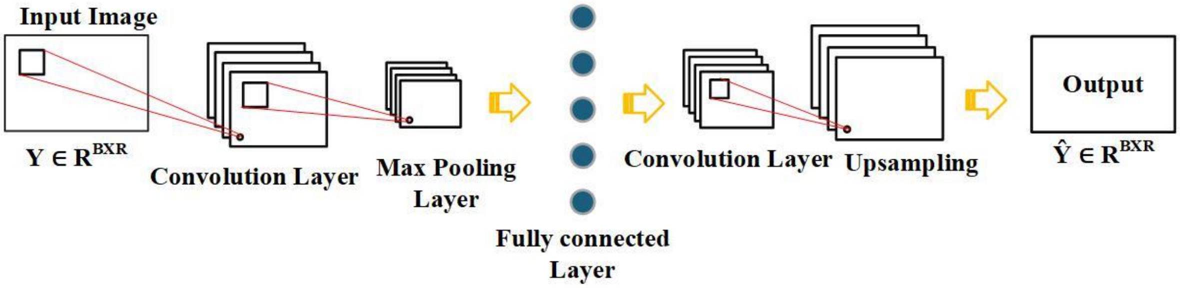


Fig. 2. Architecture of CNN Autoencoder.

The fitness of the individual is assessed by the objective function $f(Y_i)$based on the problem-specific goals. The measure of an individual's performance in relation to the optimization criteria is represented by this function.

The likelihood of choosing a person for development is directly correlated with their level of fitness. Proportional selection is a popular technique of selection in which candidates who are more fit have a larger probability of being chosen as shown in Eq. (5).

$$\text{Q}\ (Selection_i) = \frac{f(Y_i)}{\sum_{j=1}^{N} f(Y_j)} \qquad (5)$$

Crossover generates offspring by fusing the genetic material of two parent solutions. Genetic material is exchanged between parents to form the offspring, and the crossover point is randomly selected as in Eq. (6).

$$Offspring_h = (y_{h1}, y_{h2} \dots, y_{hi}, \dots, y_{hn}) \qquad (6)$$

A mutation modifies a person randomly; usually, this is done by flipping binary representation bits as shown in Eq. (7). The possibility of a mutation happening is determined by the mutation probability, or $Q_{mutation}$.

$$Mutation_h = (y_{h1}, y_{h2} \dots, \tilde{y}_{hi}, \dots, y_{hn}) \qquad (7)$$

where, $\tilde{y}_{hi}$ is the mutated bit.

The genetic procedures result in the creation of a new population. To guarantee that the best answers are kept, the members of the new population replace the members of the old population. Until a termination requirement is satisfied, the algorithm iterates continuously. A maximum number of generations, reaching a particular fitness level, or convergence are examples of common requirements. Through these iterations, genetic algorithms gradually evolve the population in the direction of ideal solutions. The particulars and parameters, including crossover and mutation rates, vary depending on the nature of the optimization task at hand and should be adjusted accordingly. The algorithm for CNN Autoencoder-GA is given below:

Algorithm 1: CNN Autoencoder-GA

Import the required data
Describe the CNN Autoencoder architecture
Develop the CNN Autoencoder
Specify the goal function that the genetic algorithm will optimize
Analyze the system's performance using the specified configuration
Provide a fitness score based on availability and additional pertinent data
Set the Genetic Algorithm up initially
Utilize the Genetic Algorithm to maximize system settings
For increased availability, apply the optimized configuration to the production system

V. Results and Discussion

Through the use of a data-driven approach, this study presents a unique way for improving production system availability and failure detection. The suggested architecture uses CNN Autoencoder for accurate failure detection and makes use of state-of-the-art techniques like deep learning and evolutionary algorithms. Overall availability is increased by using genetic algorithms to optimize system settings. The model recognizes intricate patterns after being trained on a variety of datasets, and evolutionary techniques guarantee system configuration flexibility. The model, which is implemented in Python, highlights the effectiveness of the data-driven approach in identifying and averting issues, reducing downtime, and optimizing production efficiency.

A. Model Accuracy

The degree of agreement between the deep learning model's predictions and the actual results in terms of failure detection and availability enhancement in a production system is known as model accuracy. It measures the model's efficiency in accurately detecting malfunctions and maximizing system availability. The accuracy statistic measures the percentage of cases that are properly categorized out of all the instances that the model evaluates. A high model accuracy suggests a stable and dependable performance, demonstrating the effectiveness of the data-driven, deep learning approach in conjunction with genetic algorithms in accurately predicting failure events and improving production system availability overall.

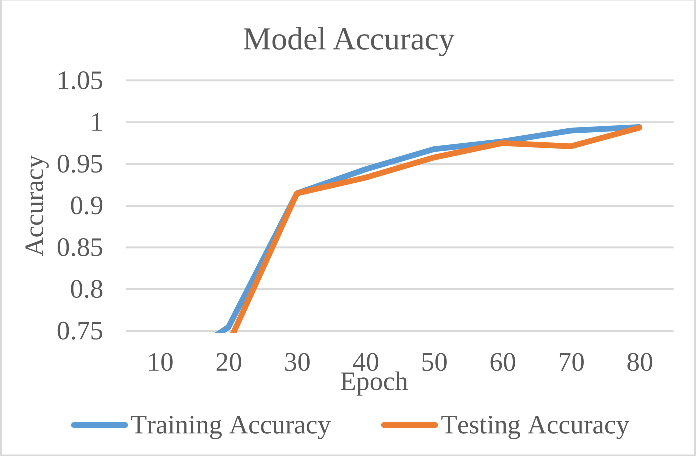


Fig. 3. Model accuracy.

The visual model accuracy graph illustrated in Fig. 3 how well the deep learning-based, data-driven strategy performed in improving production system performance. Any upward trend on the accuracy graph demonstrates an increase in the model's capacity to accurately forecast failure occurrences and raise system availability. In the context of failure detection and availability enhancement, fluctuations or plateaus may indicate regions that require extra data collection or more optimization to reach greater levels of accuracy.

B. Model Loss

When it comes to failure detection and availability enhancement in a production system, model loss is the quantitative measure of how different the actual observed values are from the projected outcomes produced by the deep learning model. This measure captures the discrepancy between the model's predictions and the actual data, indicating the degree of departure or mistake in the model's predictions. One of the main goals of training procedures is to reduce the model loss, which indicates how well the model is able to capture and reflect the underlying patterns in the data. In the context of this study, minimizing model loss is essential to guaranteeing the efficacy of the data-driven, deep learning technique and the use of genetic algorithms for production system performance optimization via improving availability and failure detection.

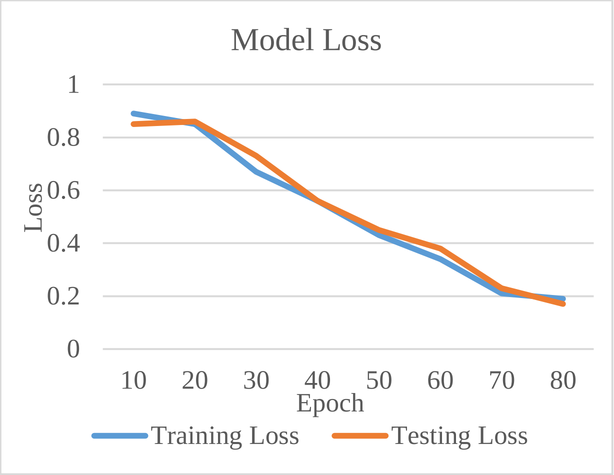


Fig. 4. Model loss.

The deep learning model's predicted mistakes are shown to have evolved in the model loss graph in Fig. 4 within the framework of improving production system performance. The model's enhanced capacity to reduce differences between expected and actual results, especially in failure detection and availability augmentation, is indicated by a declining trend in the loss graph. Finding trends or plateaus in the loss graph could inspire additional research into improving the model's architecture or training parameters to provide predictions that are more accurate.

C. ROC Curve

A graphical depiction known as the Receiver Operating Characteristic (ROC) evaluates the deep learning model's performance in terms of the trade-off between true positive rates and false positive rates.

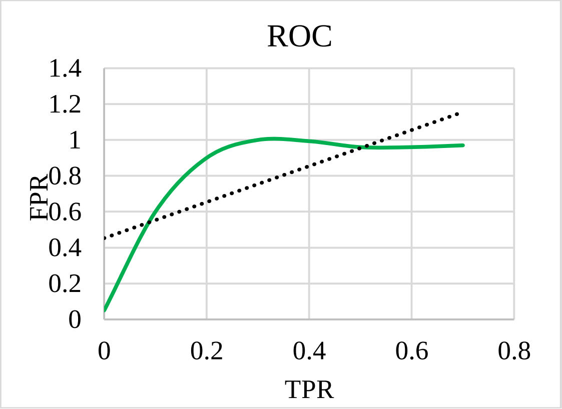


Fig. 5. ROC.

As it is depicted in Fig. 5, a thorough understanding of the model's capacity to distinguish between positive and negative occurrences linked to failure detection and availability enhancement in a production system is offered by the ROC curve, which plots the sensitivity (true positive rate) against 1-specificity (false positive rate) at various threshold settings. An increased area under the ROC curve signifies enhanced model performance in terms of accurate prediction-making while accounting for the proper ratio of true positives to false positives.

D. Performance Metrics

1) Accuracy: Accuracy is used to evaluate the system model's overall performance. Its fundamental premise is that all interactions are foreseeable. The accuracy is provided by Eq. (8).

$$Accuracy = \frac{T_{Pos} + T_{Neg}}{T_{Pos} + T_{Neg} + F_{Pos} + F_{Neg}} \tag{8}$$

2) Precision: Precision describes how comparable two or more calculations are to each other in addition to being correct. The link between accuracy and precision shows how quickly opinions may change. It is discussed in Eq. (9).

$$P = \frac{T_{Pos}}{T_{Pos} + F_{Pos}} \tag{9}$$

3) Recall: The percentage of all pertinent discoveries that were correctly sorted utilizing the procedures is known as recall. By dividing the genuine positive by the erroneously negative values, one may get the proper positive for these integers. The passage is located in Eq. (10).

$$R = \frac{T_{Pos}}{T_{Pos} + F_{Neg}} \tag{10}$$

4) F1-Score: The F1-Score computation combines recall and accuracy. To find the F1-Score, use (11), this divides the recall by the accuracy.

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{11}$$

TABLE. I MODEL PERFORMANCE

Model Performance	Percentage (%)
Accuracy	95.54
Precision	93.78
Recall	98.67
F1-Score	99.32

The model given has great efficacy in anomaly identification, as indicated by the performance indicators, which are summarized in Table I. With an accuracy of 95.54%, the model's predictions are shown to be generally accurate. With a precision of 93.78%, the model is able to correctly detect true positives among the occurrences that it classifies as anomalies. At 98.67%, the recall rate is quite high, indicating that the model is capable of accurately identifying a significant proportion of real abnormalities. Moreover, the F1-Score a balanced statistic that takes recall and accuracy into account stands out at 99.32%, highlighting the model's resilience in reaching a pleasing combination of recall and precision. Together, these measures highlight the model's excellent performance in industrial systems' real-time anomaly detection, highlighting its capacity for precise identification and reduction of false positives and false negatives.

TABLE. II COMPARISON OF PERFORMANCE METRICS

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN-LSTM [26]	95.54	94.67	94.55	92.44
Bi-LSTM [27]	93.78	97.12	92.63	95.87
CNN-RNN [28]	98.67	97.35	94.69	94.26
Proposed CNN Autoencoder-GA	99.32	99.12	98.99	98.34

Table II shows the success of various techniques in the enhancing production system performance research is demonstrated by the classification performance metrics that are supplied. With an astounding accuracy of 99.32%, the Proposed CNN Autoencoder-GA notably surpasses existing models, demonstrating its capacity to accurately identify occurrences relevant to availability enhancement and failure detection.

Furthermore, a significant percentage of real positive cases are captured by the model, as seen by the high recall (98.99%) and precision (99.12%) values. The robustness of the suggested model is further shown by the F1-Score of 98.34%, which takes into account the harmonic mean of accuracy and recall. By comparison, the CNN-RNN approach has a high accuracy of 98.67%, highlighting its ability to make accurate predictions. These thorough metrics offer insightful information on the advantages of each model, with the Proposed CNN Autoencoder-GA demonstrating significant promise as a method for obtaining better results in the intended production system applications. It is depicted in Fig. 6.

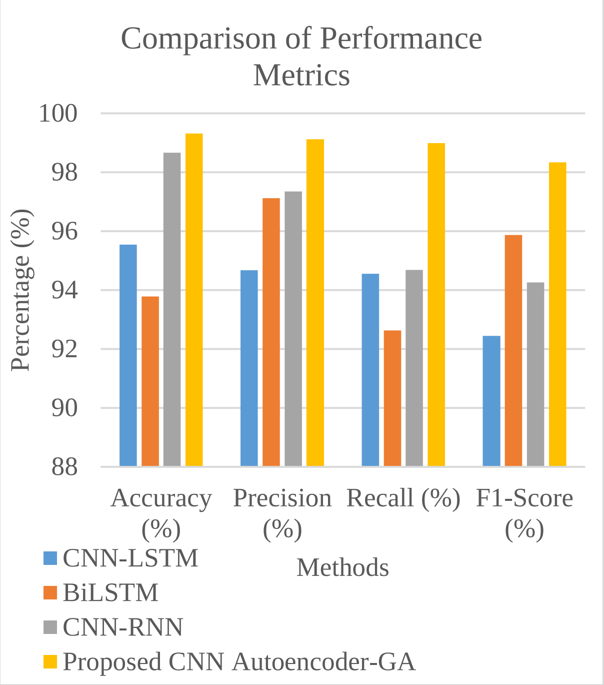


Fig. 6. Comparison of performance metrics.

E. Discussion

Through the integration of state-of-the-art methods for failure detection and availability enhancement in production systems, the suggested approach leads the way in the advancement of industrial operations. Through the use of CNN Autoencoder and deep learning, the model is able to identify complex patterns in a variety of datasets that contain data on equipment failures. A proactive approach to failure identification is made possible by this strong feature extraction capacity, which enables preventive measures to alleviate possible problems before they worsen. Furthermore, system configurations are optimized through the use of genetic algorithms, improving overall availability. With its ability to adapt to changing operating circumstances, this adaptive optimization mechanism provides a robust and effective production system management solution.

The suggested method stands out because to its exceptional accuracy of 99.32%, outperforming state-of-the-art methods such as CNN-LSTM[26], BiLSTM [27], and CNN-RNN [28]by 3.58%. This exceptional accuracy is ascribed to the complementary work of genetic algorithms and deep learning, which together offer a comprehensive approach to availability enhancement and failure detection. The suggested model, in contrast to traditional techniques, captures intricate linkages in the data, allowing for a more sophisticated comprehension of possible errors. By guaranteeing flexibility and resilience in system configurations, the application of genetic algorithms further sets the technique apart and addresses the inherent difficulties of industrial production processes. All things considered, the suggested methodology represents a major breakthrough in failure detection and availability enhancement, not only surpassing previous approaches but also providing a proactive and adaptable solution to dynamic difficulties in industrial situations.

Methodological relevance and industry prevalence served as the foundation for the benchmarking approaches chosen for this investigation. The selection of CNN-LSTM, Bi-LSTM, and CNN-RNN as typical benchmarks for well-established approaches stemmed from their extensive use in industrial settings for time-series data processing. LSTM variations such as CNN-LSTM and Bi-LSTM, which concentrate on managing temporal dependencies, tackle the sequential character of industrial data and guarantee a thorough assessment of the temporal relationship capabilities of the suggested model. Furthermore, the incorporation of CNN-based techniques takes into account the intricacy of equipment failure data, enabling a thorough evaluation of the suggested model's capacity to identify complex spatial patterns. This comprehensive comparison, which includes both CNN-based and LSTM-based approaches, offers a modern, industry-relevant framework that guarantees a careful assessment of the suggested deep learning and genetic algorithm strategy in the context of enhancing production system performance.

VI. Conclusion and Future Scope

This research has shown encouraging outcomes for improving availability and detecting failures, with a focus on using genetic algorithms. With its impressive 99.32% accuracy as well as its excellent precision, recall, and F1-Score values, the suggested CNN Autoencoder-GA model stands out as a reliable option for handling the difficulties associated with complicated production systems. The model's stability and effectiveness are further enhanced by the preprocessing stage's use of Min-Max Normalization, which guarantees that the model can absorb and analyze a variety of standardized input data. The study emphasizes how important cutting-edge methods like CNN and Autoencoder are for identifying complex patterns in real-world data, and how working in tandem with genetic algorithms improves the model's capacity to optimize for better availability and failure detection. In the future, this study will focus on extending the suggested approach's utility to other industrial contexts and investigating how well it can be tailored to real-time production scenarios. Furthermore, additional research on the interpretability of the model's decision-making procedures would improve industry acceptance and confidence in the suggested technique. The model's performance may be further enhanced by including more complex evolutionary algorithms or investigating hybrid models that incorporate other optimization methods. Furthermore, in order to support large-scale production systems, the scalability of the deep learning-based technique should be investigated. The landscape of intelligent systems for fault detection and availability enhancement in industrial settings will be significantly shaped by the ongoing improvement and adaption of these approaches as technological developments persist. All things considered, this work establishes the groundwork for a proactive and effective strategy for managing production systems by utilizing genetic algorithms in conjunction with deep learning.

The study's efficacy may be impacted by real-world unpredictability, and its generalizability to various industrial settings and datasets is restricted. The computational resources required for optimizing system setups and training the deep learning model may give rise to practical limits. The emphasis on accuracy measures obscures a thorough evaluation of the robustness of the model in different scenarios or with respect to outside influences. Furthermore, the Python software implementation might make it more difficult to integrate seamlessly with production systems that use other technology stacks.

References

[1] A. Essien and C. Giannetti, ‘A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders’, IEEE Transactions on Industrial Informatics, vol. 16, no. 9, pp. 6069–6078, Sep. 2020, doi: 10.1109/TII.2020.2967556.

[2] W. Sun, J. Liu, and Y. Yue, ‘AI-Enhanced Offloading in Edge Computing: When Machine Learning Meets Industrial IoT’, IEEE Network, vol. 33, no. 5, pp. 68–74, Sep. 2019, doi: 10.1109/MNET.001.1800510.

[3] H. Wang, S. Li, L. Song, L. Cui, and P. Wang, ‘An Enhanced Intelligent Diagnosis Method Based on Multi-Sensor Image Fusion via Improved Deep Learning Network’, IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 6, pp. 2648–2657, Jun. 2020, doi: 10.1109/TIM.2019.2928346.

[4] J. P. Yun, W. C. Shin, G. Koo, M. S. Kim, C. Lee, and S. J. Lee, ‘Automated defect inspection system for metal surfaces based on deep learning and data augmentation’, Journal of Manufacturing Systems, vol. 55, pp. 317–324, Apr. 2020, doi: 10.1016/j.jmsy.2020.03.009.

[5] Q. Wang, W. Jiao, and Y. Zhang, ‘Deep learning-empowered digital twin for visualized weld joint growth monitoring and penetration control’, Journal of Manufacturing Systems, vol. 57, pp. 429–439, Oct. 2020, doi: 10.1016/j.jmsy.2020.10.002.

[6] M. Zhang, F. Tao, and A. Y. C. Nee, ‘Digital Twin Enhanced Dynamic Job-Shop Scheduling’, Journal of Manufacturing Systems, vol. 58, pp. 146–156, Jan. 2021, doi: 10.1016/j.jmsy.2020.04.008.

[7] K. Alexopoulos, N. Nikolakis, and G. Chryssolouris, ‘Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing’, International Journal of Computer Integrated Manufacturing, vol. 33, no. 5, pp. 429–439, May 2020, doi: 10.1080/0951192X.2020.1747642.

[8] ‘Electronics | Free Full-Text | Artificial Intelligence-Based Decision-Making Algorithms, Internet of Things Sensing Networks, and Deep Learning-Assisted Smart Process Management in Cyber-Physical Production Systems’. Accessed: Nov. 23, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/10/20/2497

[9] M. Xia, H. Shao, D. Williams, S. Lu, L. Shu, and C. W. de Silva, ‘Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning’, Reliability Engineering & System Safety, vol. 215, p. 107938, Nov. 2021, doi: 10.1016/j.ress.2021.107938.

[10] R. Rai, M. K. Tiwari, D. Ivanov, and A. Dolgui, ‘Machine learning in manufacturing and industry 4.0 applications’, International Journal of Production Research, vol. 59, no. 16, pp. 4773–4778, Aug. 2021, doi: 10.1080/00207543.2021.1956675.

[11] Y. Li, X. Du, F. Wan, X. Wang, and H. Yu, ‘Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging’, Chinese Journal of Aeronautics, vol. 33, no. 2, pp. 427–438, Feb. 2020, doi: 10.1016/j.cja.2019.08.014.

[12] ‘Sensors | Free Full-Text | Bearing Fault Diagnosis of Induction Motors Using a Genetic Algorithm and Machine Learning Classifiers’. Accessed: Nov. 23, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/20/7/1884

[13] J. Gao, H. Wang, and H. Shen, ‘Task Failure Prediction in Cloud Data Centers Using Deep Learning’, IEEE Transactions on Services Computing, vol. 15, no. 3, pp. 1411–1422, May 2022, doi: 10.1109/TSC.2020.2993728.

[14] ‘WaveletKernelNet: An Interpretable Deep Neural Network for Industrial Intelligent Diagnosis | IEEE Journals & Magazine | IEEE Xplore’. Accessed: Nov. 23, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9328876

[15] L. Sun and F. You, ‘Machine Learning and Data-Driven Techniques for the Control of Smart Power Generation Systems: An Uncertainty Handling Perspective’, Engineering, vol. 7, no. 9, pp. 1239–1247, Sep. 2021, doi: 10.1016/j.eng.2021.04.020.

[16] B. Farooq, J. Bao, J. Li, T. Liu, and S. Yin, ‘Data-Driven Predictive Maintenance Approach for Spinning Cyber-Physical Production System’, J. Shanghai Jiaotong Univ. (Sci.), vol. 25, no. 4, pp. 453–462, Aug. 2020, doi: 10.1007/s12204-020-2178-z.

[17] M.-A. Filz, J. E. B. Langner, C. Herrmann, and S. Thiede, ‘Data-driven failure mode and effect analysis (FMEA) to enhance maintenance planning’, Computers in Industry, vol. 129, p. 103451, Aug. 2021, doi: 10.1016/j.compind.2021.103451.

[18] H. Chowdhury, ‘Semiconductor Manufacturing Process Improvement Using Data-Driven Methodologies’. Preprints, Oct. 07, 2023. doi: 10.20944/preprints202310.0056.v2.

[19] A. Kara, ‘A data-driven approach based on deep neural networks for lithium-ion battery prognostics’, Neural Comput & Applic, vol. 33, no. 20, pp. 13525–13538, Oct. 2021, doi: 10.1007/s00521-021-05976-x.

[20] M. Ghahramani, Y. Qiao, M. C. Zhou, A. O’Hagan, and J. Sweeney, ‘AI-based modeling and data-driven evaluation for smart manufacturing processes’, IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 4, pp. 1026–1037, Jul. 2020, doi: 10.1109/JAS.2020.1003114.

[21] S.-K. S. Fan, W.-K. Lin, and C.-H. Jen, ‘Data-driven optimization of accessory combinations for final testing processes in semiconductor manufacturing’, Journal of Manufacturing Systems, vol. 63, pp. 275–287, Apr. 2022, doi: 10.1016/j.jmsy.2022.03.014.

[22] J. Friederich and S. Lazarova-Molnar, ‘Towards Data-Driven Reliability Modeling for Cyber-Physical Production Systems’, Procedia Computer Science, vol. 184, pp. 589–596, Jan. 2021, doi: 10.1016/j.procs.2021.03.073.

[23] J. Wu, K. Hu, Y. Cheng, H. Zhu, X. Shao, and Y. Wang, ‘Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network’, ISA Transactions, vol. 97, pp. 241–250, Feb. 2020, doi: 10.1016/j.isatra.2019.07.004.

[24] A. Choudhary, T. Mian, and S. Fatima, ‘Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images’, Measurement, vol. 176, p. 109196, May 2021, doi: 10.1016/j.measurement.2021.109196.

[25] ‘Machine Failure Prediction | Kaggle’. Accessed: Nov. 21, 2023. [Online]. Available: https://www.kaggle.com/code/bhaveshjain1612/machine-failure-prediction

[26] ‘Application of CNN-LSTM in Gradual Changing Fault Diagnosis of Rod Pumping System’. Accessed: Nov. 23, 2023. [Online]. Available: https://www.hindawi.com/journals/mpe/2019/4203821/

[27] ‘DeepRan: Attention-based BiLSTM and CRF for Ransomware Early Detection and Classification | Information Systems Frontiers’. Accessed: Nov. 23, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s10796-020-10017-4

[28] ‘Acoustic Anomaly Detection of Mechanical Failure: Time-Distributed CNN-RNN Deep Learning Models | SpringerLink’. Accessed: Nov. 23, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-19-3923-5_57.

Enhanced Multi-Object Detection via the Integration of PSO, Kalman Filtering, and CNN Compressive Sensing

S. V. Suresh Babu Matla[1], Dr.S. Ravi[2*], Muralikrishna Puttagunta[3]
Research Scholar, Department of Computer Science-School of Engineering and Technology, Pondicherry University, India[1]
Professor, Department of Computer Science-School of Engineering and Technology, Pondicherry University, India[2]
Assistant professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, 522502, Andhra Pradesh, India[3]

***Abstract*—Many inventive techniques have been created in the field of machine vision to solve the challenging challenge of detecting and tracking one or more objects in the face of challenging conditions, such as obstacles, object motion, changes in light, shaking, and rotations. This research article provides a novel method that combines Convolutional Neural Networks (CNNs), Compressive Sensing, Kalman Filtering, and Particle Swarm Optimization (PSO) to address the challenges of multi-object tracking under dynamic conditions. Initially, a CNN-based object classification and identification system is demonstrated, which efficiently locates objects in video frames. Subsequently, in order to produce precise representations of object appearances, utilizing compressive sensing techniques. The Kalman Filter ensures adaptability to irregular observations, eliminates erroneous data, and reduces uncertainty. PSO enhances tracking efficiency by optimizing forecast precision. When combined, these techniques provide robust tracking even in the presence of complex movement patterns, occlusions, and visual disparities. The efficiency of this strategy is demonstrated by an empirical investigation that produces a remarkable tracking accuracy of 98%, which is 3.15% greater than other methods across a range of challenging settings. This technique has been compared to various existing approaches, including the Clustering Method, YOLOV4 DNN Model, and YOLOV3 Model, and its deployment is made easier with Python software. This hybrid technique, which addresses the limitations of separate approaches and offers a holistic approach to multi-object monitoring, has potential applications in surveillance, robotics, and autonomous systems.**

Keywords—Multi-Object tracking; object detection; convolutional neural networks; kalman filtering; particle swarm optimization

I. INTRODUCTION

Computer vision is a dynamic investigation area, which widens its perceptual field across a range of purposes like traffic and navigation systems, motion detection, speech and voice recognition and other applications. In which Object detection by means of a single aspect of the target will not provide a favorable resolution, as the object manifestation may altered depending on light effects, object position shifts, and blurring effects, occlusion and so on. As a result, multi-scale object detection was planned, which delivers further healthy resolution than single-degree object detection [1]. Quite a lot of discriminative appearance Random projections have been expected to venture high dimensions properties like features to a low dimensional space. This compressive sensing is self-sufficient in terms of data, as well as non-adaptive naïve Bayes classifier is used to differentiate among optimistic and undesirable data [2]. Extracts the features through compressive sensing and sparse representation to construct the model more fast and robust, and applied the LS-SVM classifier act as the optimization algorithm to separate positive templates from negative samples, finally hypergraph methodology use to improve the tracking accurately [3].

Majority of the object detection techniques are restricted to only a limited object group, the classification of datasets and of images will be a monotonous task. A training procedure called YOLO9000 was projected, This is a real-time object detector capable of detecting numerous items, including over 9000 samples [4]. In order to detects multiple objects, a Kalman filtering is secondhand, whose object structures will be arbitrary and Gaussian in fauna. Gaussian in personality. Once the location of the item transforms regularly and the movement is piercing, the Kalman filter is secondhand. Kalman filtering has two stages: 1. Prediction, in which the object location is predicted; and 2. Correction, in which the estimated value is approved in relation to the predetermined state assessment. The Particle Swarm Optimization (PSO) method is an extremely efficient object tracking technique that follows objects in moving video in a manner akin to bird flocking. The ideal object frame may be found based on the item's current location, its prior best position, and its best position in the overall frame.

Sparse representation object model is the majority and most commonly used in object recognition, two types of imaging sets were observed to apply. One is for object patterns, which have pixel values from previous frames. Another is inconsequential templates, which include noisy pixels. The pixels that make up object patterns are sparse manner described in a low-dimensional sub space, resulting in a path of non-zero entrances in the sparse medium, which represents the location of the pixel in a precise image frame [5]. However, the object model with thin depiction grieves from heavy obstruction and cannot account for things that come and disappear over time.

Object tracking in video clips is a problem for the available object tracking techniques. For applications ranging from surveillance to robotics and autonomous systems, the capacity to reliably detect and monitor objects in the midst of barriers, changing illumination conditions, object motion, shaking, and rotations is essential. Occlusions, scale shifts, and visual alterations are some of these difficulties. To accomplish this, need to overcome the above-mentioned issues, the paper has proposed CNN based FCT (Fast Compressive Tracking) method in which the sparse random matrix generated multiple object features is fed to the CNN model. The input is fed as two successive image frames, as the predominant features are extracted through the processing of stack of CNN layers and fed to the output layer. In order to improve object representation accuracy, compressive sensing techniques are included. The purpose of the Kalman Filter is to reduce uncertainty, remove erroneous data, and provide adaptation to irregular observations. At the output of the CNN model, Kalman filtering and Particle Swarm Optimization (PSO) is applied in order to locate the locations of the target image features. Particle Swarm Optimization is included to maximize forecasting accuracy and increase tracking efficiency. When these methods are combined, a strong solution that can handle problems like intricate movement patterns, occlusions, and visual imbalances is produced. A Support Vector Machine (SVM) classifier is trained in order to classify the target features of the images from the rest of the images. The proposed algorithm was able to track and detect objects under severe occlusion as well as was able to track images which are inconsistent in nature. CNN used extensively in the area of visual object tracking, with a strong ability to learn distinctive image features depictions, feature map selection methods to select discriminative features and discard noisy or unrelated ones. A multi-domain learning framework, called as MDNet, full network trained offline, and the connected layers counting Online fine-tuning of a single domain-specific layer [26]. CNN focused on visual recognition tasks, contains domain adaptation, fine-grained based recognition, and largescale scene recognition. The potentiality of a visual recognition system to attain elevated classification accuracy on exercise with sparse labeled data has shown to be a long term objective in computer vision research [6].

More reliable and effective multi-object identification techniques are desperately needed in dynamic and complicated situations, which is what inspired the proposed work. Situations where things are obscured, arrive suddenly, or vanish are difficult for traditional object detection techniques to handle. The CNN-FCT Methodology is selected as a ground-breaking solution in response to these constraints because of its exceptional characteristics and capabilities. To solve the issues with previous approaches, the CNN-FCT Methodology combines Convolutional Neural Networks (CNN), Particle Swarm Optimization (PSO), and Kalman Filtering. The model can recognize intricate patterns and representations in a variety of circumstances thanks to the efficient feature extraction provided by CNN. This is especially important to overcome occlusion-related difficulties, when standard approaches might not be able to recognize objects with enough accuracy. Dynamic optimization and tracking techniques are introduced via the combination of PSO and Kalman Filtering. PSO improves the model's ability to adjust to abrupt changes in the scene by improving the predicted positions of objects at each iteration. This is enhanced by Kalman Filtering, which offers a predictive filtering process that makes tracking more accurate particularly when objects appear and disappear. The shortcomings of conventional object detection techniques have been exposed by recent developments in computer vision and machine learning, particularly in situations that are extremely dynamic and unexpected. These issues are acknowledged in the suggested study, which creatively integrates CNN Compressive Sensing, PSO, and Kalman Filtering into the CNN-FCT Methodology. By doing this, the research hopes to greatly improve the performance of multi-object detection systems, strengthening their ability to withstand the intricacies of the real world and advancing computer vision applications across a range of industries.

The key contribution of the paper is given as follows:

- The proposed approach differentiates out since it integrates several innovative techniques. The system integrates CNNs, Compressive Sensing, Kalman Filter integration, and PSO to enhance object tracking effectiveness.

- An effective and precise item categorization and identification system is facilitated by the methodology's early integration of Convolutional Neural Networks (CNNs). This section makes sure that objects are located precisely inside of video frames, providing a solid basis for tracking operations.

- The accuracy of object representations is further improved by the application of compressive sensing methods. This addition improves object representation accuracy by effectively extracting relevant information from the video frames.

- By removing false data and lowering uncertainty in the tracking process, the inclusion of Kalman Filtering gives flexibility to erratic observations. By allowing the technique to adjust to erratic variations in object movement, this contribution enhances the methodology's durability and offers a more precise and dependable tracking mechanism.

- Particle Swarm Optimization is included to improve forecasting accuracy and tracking efficiency. This optimization process helps to maintain object tracking accuracy even in scenarios with occlusions, appearance modifications, and complicated motion patterns.

The remaining sections of the paper are ordered as follows: Section II deliberates about various multi-object tracking algorithms, Section III portrays about problem statement, Section IV discloses about the kernel filtering and sparse representation model, Section V gives a detailed explanation about the proposed CNN-FCT model, and Section VI depicts the investigational consequences and accomplishes the paper.

II. Related Works

As evidenced by this literature review, there are several visual object tracking methods available, we are going to assess few algorithms and methodologies, which have been extensively used for exploring reason.

A. *Fast Visual Tracking*

Li and Wang has proposed Dense Spatio-temporal Context Learning for fast graphic tracking [7]. This method creates a Bayesian framework, for associating the spatio-temporal aspects among the goal object and the foreground and circumstantial images pertaining to the image frames. For rapid learning and detection, this work applied four- Fast Fourier Transformation (FFT) and the results showed that this framework performed well against state-of-the-art techniques with admiration to effectiveness, correctness and robustness.

Bern et, al. [8] has proposed a incremental model algorithm, in which the target obejct is tracked increamently to a low-dimensional subspace presentation and also adapts to the dynamic changes pertaining to the target object. The increamenting algorithm is based on the principal component analysis and forgetting factor, that helps in enhancing the overall performance of the proposed tracking algorithm.

Najeeb and Ghani has made comparative study on the techniques of object tracking in various applications pertaining to computer vision domain namely, traffic surveillance, robot navigation and so on [9]. This paper studied about various object tracking methods, which tracks single or multiple objects in motion form a video sequence. The major object tracking techniques discussed in this survey are kernel tracking, point tracking and Silhouette tracking algorithms.

Dr. D. S. David [10] has proposed innovative and efficient object detection and tracking algorithm that utilizes optical flow in combination with motion vector assessment for object discovery and trailing in a series of frames. The optical flow exposes the details of the moving object and motion vector assessment provides the position of an object from successive frames and helps in increased accuracy in spite of motion blurs and cluttered image background.

B. *Particle Swarm Optimization (PSO)*

Moussa and Shoitan in this work has implemented Sequential Particle Swarm Optimization (PSO) for visual tracking, which tracks the objects by including significant features of the tracked objects [11]. This method resulted in increasing tracking performance because of its parameters that is flexible according to the fitness value of the objects and predicts the object's location correctly when it is in motion.

Nedjah has proposed an object tracking algorithm, which utilizes PSO technique, in order to detect the target object in motion from a video image series [12]. The entire video sequence is searched as an individual swarms which provides optimal solution.

C. *Compressive Visual Tracking*

Chen et al. [13] has proposed a scheme along with weight maps called as multi-sparse measurement matrix, which compresses the image but at the same time maintains the originality of the image features. The weight map merges a distinct weight as well as a characteristic weight to proficiently differentiate the aimed manifestation and position. Furthermore, a dispersion function is applied for the digital updating of the intended template letting to track together the position and extent of the target object.

Nguyen et al. [14] proposed a method which holds two steps one is effective Cholesky decomposition on GPU and FPGA. In order to improve the performance with respect the memory access, second is CS signal reconstruction applied on FPGAs and GPUs which helps to reduce the computation.

Li et al. [15] has proposed template matching dealing out phase along with the compressive tracking method, in demand to maintain the constant frame frequency and rise the strength of the tracker. This template comprises of all the available data and similarity metrics have been used to compare the available data with the regularly used images series.

D. *Kalman Filtering*

Kaderali in this work has implemented IMM cubature Kalman filter (IMM-CKF) in order to track the orbiting space objects [16]. This study utilizes the geometric association among the planetary object specifically space craft, space based optical (SBO) sensor, and the sun for tracking the space object. A situation which encompasses the aimed spacecraft and four SBO sensors is utilized to check efficiency of the IMM-CKF. A comparison was made between IMM-CKF and the normal cubature Kalman filter (CKF). The outcome designates that IMM-CKF is extremely vigorous than the CKF when the space object experiences a movement. Shantaiya et al. [5] has proposed multiple objects tracking from the video series using optical flow algorithm. This algorithm uses Kalman filtering to detect and track the moving objects in each frame, which helps in identifying moving objects with occlusion, blurred objects and so on. Ullah et al. [17] for the usage against non-linear systems proposed a novel method known as Unscented Extended Kalman Filtering (UEKF). This new method of Kalman filtering is same as traditional Kalman filtering except, for every running non-linear sample, the deterministic sample is caught with non-linear mean and linear covariance. This new method offered better performance when compared to that of conventional Kalman filtering.

E. *Convolutional Neural Network-based Visual Tracking*

Xiao and Pan has researched, how CNN filers helps in tracking moving objects by increasing the depth of the CNN [20]. It was found that, by making the depth to 16-19 layers, the proposed ConvNet showed better performance than the traditional visual tracking methodologies when evaluated against state-of-arts results. Mahmoudi et al. [18] has proposed CNN based online moving object tracking algorithm, in which the tracking missed error rates are reduced by employing truncated structure loss function and temporal selection mechanism for discriminating positive and negative samples. Liu et al. [19] has proposed online CNN model for tracking visual objects. For hierarchical feature learning, and to reduce the dependability of CNN, k-Means is applied. Regression model is involved to notice the positive samples of the goal visual object. The investigational outcomes showed that, the planned model executed very well when compared to another Online CNN model. Li and Yang proposed a model based on rich feature hierarchies of CNNs cultured from a large-scale datasets and train a liner correlation filter on each convolutional layer and conclude the goal location with a coarse-to-fine searching method [20], yields outcomes favorably against the state-of-the-art approaches in relations of robustness and accuracy.

Object tracking techniques in computer vision, Incremental Model Algorithms, Dense Spatio-temporal Context Learning, PSO for visual tracking, Compressive Visual Tracking with weight maps and efficient signal reconstruction, Kalman Filtering for tracking orbiting space objects, and CNN-based Visual Tracking are just a few of the advanced techniques for visual tracking that are examined in this literature review. To improve tracking performance, these approaches use a variety of strategies, including CNN filters, optical flow, Bayesian frameworks, Principal Component Analysis, and Cholesky decomposition. The review illustrates their uses in various contexts and shows how well they work to increase accuracy and resilience when compared to state-of-the-art outcomes and conventional tracking approaches.

Numerous strategies have been investigated by researchers, such as incremental model algorithms, Bayesian frameworks, comparative examinations of tracking methods, novel object identification and tracking algorithms, and the use of optimization strategies like Particle Swarm Optimization (PSO). Also, there is an emphasis on improving tracking performance by utilizing deep learning techniques, namely CNNs, multi-sparse measurement matrices, compressive tracking, and Cholesky decomposition. The main difficulty, despite the variety of approaches, is creating dependable and effective object tracking systems that can deal with things like motion blur, crowded backgrounds, and shifting object appearance.

III. Problem Statement

From the above discussed literatures, it is found that the challenge involves tackling the difficulty of multi-object tracking in dynamic environments, where a large number of objects must be reliably and precisely tracked despite occlusions, visual discrepancies, and intricate movement patterns [21]. The goal is to improve tracking accuracy and resilience by combining CNNs, PSO, Kalman Filtering, and Compressive Sensing. The issue statement's primary objective is to develop and evaluate an improved multi-object tracking system that employs these techniques in order to progress object tracking and computer vision applications and achieve high tracking accuracy and reliability in challenging circumstances. The method used in the study is called Multiple Object Tracking, or MOT. In order to overcome the aforementioned problems, the current study establishes a dependable and real-time tracking device that is capable of handling the difficulties associated with multiple object tracking in a variety of instances. This is achieved by using an effective compressible tracking infrastructure based on CNN models in conjunction with Kalman filtering and PSO algorithm.

IV. Proposed CNN-FCT Model for Object Tracking

The CNN-FCT Methodology combines a number of essential elements, each of which has a unique function in tackling the difficulties involved in multi-object recognition in dynamic and complex situations. The effectiveness of the methodology is largely due to its novel features, which include the use of Particle Swarm Optimization (PSO) to improve predictions, Fast Compressive Tracking (FCT) for object tracking, Segmentation with Otsu thresholding, Kalman Filtering, and Convolutional Neural Networks (CNN) for object detection and classification. This all-encompassing strategy outperforms separate approaches and has promise for robotics, autonomous systems, and surveillance. Fig. 1 depicts the proposed CNN-FCT methodology.

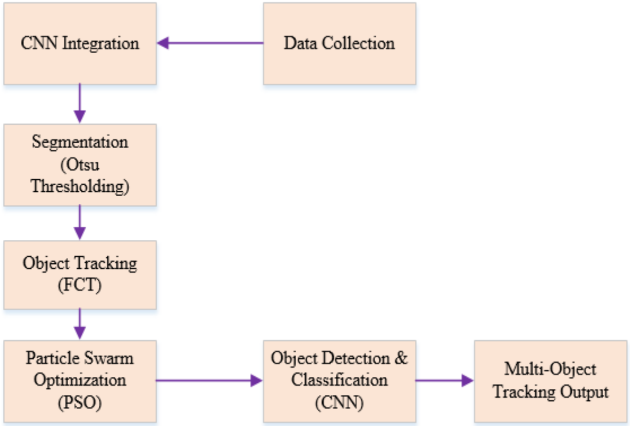

CNN Integration
Data Collection
Segmentation (Otsu Thresholding)
Object Tracking (FCT)
Particle Swarm Optimization (PSO)
Object Detection & Classification (CNN)
Multi-Object Tracking Output

Fig. 1. Proposed CNN-FCT Methodology.

A. Data Collection

The MOT challenge 2017 dataset is used for example video clips in order to evaluate the suggested approach. In all, 11245 frames are included in its 14 series of detections, seven of which are used for testing and the other seven for training [22].

B. Kalman Filtering

In the predictive tracking and filtering process, Kalman Filtering is an essential part. Kalman filtering improves the estimated item positions using dynamic measurements and predictions by implementing a recursive method. This guarantees, even in situations where objects appear, disappear, or are obscured, that their trajectories are monitored precisely. The Kalman Filter is used to improve object tracking accuracy. The Kalman Filter gives an ideal estimate of the object's state, efficiently minimizing the effects of noise and uncertainties, by integrating noisy and uncertain observations with predictive models. As fresh measurements come in, it continually updates its estimation in order to dynamically respond to changing conditions. Smooth and precise tracking is made possible by the Kalman Filter's capacity to combine historical data and forecasts with current measurements, especially in conditions characterized by noise, occlusions, and fluctuations in object speed. It is essential to attaining accurate and consistent object tracking results because of its iterative process of prediction, update, and state estimation, which acts as a strong mechanism to preserve the trajectory and location of the tracked item. Kalman filtering, which is especially effective at managing noisy data, helps to lower the number of false positives and negatives and ensure accurate estimations of the object state. The system is robust in dynamic circumstances because of its capacity to adapt to changing surroundings and retain continuity between frames. This is made possible by its sequential frame processing capabilities. By offering temporal context for object tracking when combined with CNN Compressive Sensing, Kalman Filtering enhances the capabilities of CNNs and produces a more precise, accurate, and real-time multi-object detection system that performs well in demanding and dynamic settings.

C. Segmentation

To successfully separate items from the background, the segmentation module uses Otsu thresholding. By separating foreground and background pixels with the best possible threshold, this method enhances object localization. Otsu thresholding improves the segmentation process, making it possible to extract features for later stages of the approach with greater accuracy. Otsu thresholding, a popular image segmentation approach, distinguishes between object areas and background in a video stream, which is crucial for object tracking. With the help of automation, the ideal threshold value that reduces intra-class variation in object and background areas may be found. By transforming grayscale or color frames into binary images, where pixels are classed as either object or background based on their intensity values, Otsu thresholding efficiently separates items of interest in object tracking. Otsu thresholding allows subsequent object tracking algorithms to precisely detect and follow the movements of objects over time, improving the accuracy and robustness of the tracking process. This is accomplished by segmenting the video frames into discrete zones. It is given in Eq. (1) below.

$$\sigma_z^2 = W_x\sigma_x^2 + W_y\sigma_y^2 \tag{1}$$

In this equation, W_y stands for the weight of the frontal image and σ_y^2 for its variation, whereas W_x stands for the weight of the background image and σ_x^2 for its variation. With the use of this approach, researchers may determine which pixels in an image serve to balance the foreground and background, as well as how many of them overall there are in comparison to both the background and foreground pixels. The average and variance are then calculated for the suitable backdrop and foreground. The weight and variation were then used to establish the various thresholding level values.

D. FCT for Object Tracking

For effective and real-time object tracking, Fast Compressive Tracking (FCT) integration is essential. Using compressive sensing techniques, FCT is able to track and rebuild objects over several frames. This method offers strong tracking capabilities and is especially useful in situations where objects are obscured or move suddenly. Feature correspondence and transformation techniques are used in FCT, a unique approach to object tracking, to improve the accuracy and resilience of tracking in complicated visual situations. In order to build object correspondences, this technique first extracts and matches characteristic features from successive frames. An adaptive transformation procedure then adjusts for variations in object appearance, size, and orientation. FCT is a potential approach for multi-object tracking in real-world applications because it combines feature-based matching with transformation-based modelling to address issues including occlusions, illumination changes, and object deformations. By offering a more robust and flexible framework for monitoring objects across a variety of applications, including robotics, autonomous systems, and surveillance, this method has the potential to substantially enhance the fields of computer vision and object tracking.

Fast Compressive Tracking (FCT), a ground-breaking technique for reliable and effective object tracking in video sequences, has gained prominence. FCT presents a revolutionary method for collecting and describing the look of objects, building on the concepts of compressive sensing, enabling real-time tracking even in difficult circumstances. FCT makes use of a condensed image of the object's appearance. Online learning of this condensed model is effective and adapts to changes in object appearance and background noise. The distinguishing strength of FCT is its robustness to occlusions, changes in illumination, and deformations while still successfully separating the object from the backdrop. FCT is particularly suited for high-speed applications like robotics, surveillance, and interactive systems since it can achieve amazing tracking speeds while drastically decreasing computing overhead. Compressive sensing's effective signal capture combined with FCT's adaptable appearance modelling results in a significant improvement in object tracking. Algorithm 1 explains the object tracking.

Algorithm 1: Multi-Object FCT Algorithm

```
Input: A representative model from $Y^t, Y, t_{th}$ frame.
Step 1: retrieve $v(Y^t)$ by selecting the searching range " $\kappa_y \geq 1$ and looking for step ʊs≥ 1".
Step 2: classifier CNN equations are used, the tracking location $I_{(t-1)}$ is achieved with the best predicted outcome.
Step 3: retrieve $v'(Y^t)$ by selecting the searching range " $\kappa'_y \geq 1$and $\leq \kappa_y$ and looking for step ʊ's≥ 1 and≤ ʊs".
Step 4: By using $v'(Y^t)$ to CNN equations, the tracking location $I_{(t)}$ is achieved with the best predicted outcome.
Step 5: Now, extract the features from the two sets $v\,(Y^t)$ and $v'(Y^t)$ .
Step 6: Apply the sampled feature to the SVM classifier in the SVM equation now.
Output: Object is tracked at $I_{(t)}$.
```

E. Object Detection and Classification with CNN

Utilizing Convolutional Neural Networks (CNN), the approach performs exceptionally well in object classification and feature extraction. Because CNNs are skilled at extracting hierarchical representations from incoming data, the model can recognize complex patterns and features. This improves the methodology's ability to recognize and classify objects accurately, making it appropriate for a wide range of item kinds and intricate situations. The way of recognizing and classifying things inside images and video frames has been completely transformed by the state-of-the-art computer vision. CNNs are able to identify things with great precision because of their inherent ability to automatically learn hierarchical properties from input. Using this method, a CNN model is trained on large datasets to identify particular classes or objects. The trained CNN is used to input images during the inference phase, when it searches the whole scene for objects and then gives them class labels. Numerous real-world uses for this technol8ogy exist, such as in driverless cars, spying, the interpretation of medical images, and more. Its capacity to manage intricate situations like object occlusions and changing object sizes and orientations makes it a potent instrument in the field of computer vision, opening the door for sophisticated applications in a variety of sectors.

CNNs are used for object identification and classification in computer vision applications, such as multi-object tracking and real-time situations. These applications benefit from hierarchical feature extraction, spatial hierarchies, and robustness to fluctuations. CNNs are an invaluable tool for applications requiring precise and effective object detection in a variety of visual settings because to their versatility and adaptability. In the field of object tracking, CNNs are used to locate and identify objects inside video frames, giving not only their locations but also the labels that go along with them. Even in a variety of tough visual environments, the CNN can reliably distinguish between multiple object categories thanks to its capacity to learn intricate features and patterns from large training data. CNNs are adept at classifying objects; they give labels according to traits they have learnt and are adaptable to different categories. Training is accelerated by utilizing transfer learning, particularly in situations when task-specific data is scarce. With improved designs like SSD or YOLO, CNNs find applications in real-time activities and exhibit robustness to fluctuations in object appearance.

The input is fed to the CNN network, as a combination of two cropped regions from two next frames as explained in Algorithm-2. Then it is propagated through five layers of CNN layer, where researchers remove all the remaining layers from the pooling layers to store only precise information of the cropped pair of input images. Researchers produce a couple of frames (CF_t, CF_{t-1}) with target center $\left(a+\frac{w}{2}\,;b+\frac{h}{2}\right)_{t-1}$,and size $(p_w\,, p_{sh})_{t-1}$, where, $t-1$ stands for the directory of preceding frame and p describes the exploration range.

The collected areas are resized into $m_i * m_i$ with scale factor $(sx; sy)$ which is well-defined beforehand affording with the real size of images, and nourished into a CNN model to attain the 1^{th} Convolutional beginnings. Tracking information linked list is presented to discover novel substances of attention incoming the image which is a sequence of protuberances N depicting noticed objects which emerges in frames with the same ID from the initial to the final. Each node N_t comprises the objects in CF_t. Every object has ID which is identified by its aspects. When the similar object emerges over again, it goes on the olden ID in its place of a new one. Hence, the tracking information linked list can be applied to resolve obstruction issues and to forecast misrepresented objects. The suggestion is made by the PSO algorithm.

Algorithm 2: CNN based Classification Algorithm

Input: A pair of cropped frames $(CF_t; CF_{t-1})$ target center $(a+\frac{w}{2}; b+\frac{h}{2})_{t-1}$, and
size $(p_w; p_{sh})$

Step 1: Resize the cropped pairs $m_i * m_i$ with scale factor $(s_x; s_y)$

Step 2: The resized cropped image is fed to the CNN network to obtain 1^{th} convolution activation

Step 3: The precise features are extracted from the cropped inputs by removing all the images in the pooling.

Output: Classified object linked list obtained with Node N_t and associated objects O_t with ID for every node.

F. PSO for Enhancing Predictions

To improve and refine the predictions produced by the approach, Particle Swarm Optimization (PSO) is utilized. Accuracy and flexibility are enhanced through parameter optimization of the model by PSO. Ensuring that the model dynamically adapts to changing conditions is especially useful in situations when abrupt changes take place. PSO is essential for object tracking since it improves predictions after CNN classification. Following the CNN's classification and identification of the objects in the video frames, PSO enters to improve the original predictions by maximizing the predicted object locations and trajectories. By utilizing PSO's optimization skills, the tracking algorithm may fine-tune the anticipated item positions based on elements like motion models, historical data, and real-time measurements. This improves the initial forecasts' accuracy and allows them to be adjusted to the tracking scenario's dynamic nature. In order to account for uncertainties and imperfections caused during object categorization and initial tracking, PSO's repeated optimization procedure refines the anticipated locations. Its combination with other methods, such CNN-based object recognition and Kalman Filtering, results in a hybrid strategy that provides a thorough plan for precise and adaptable forecasts in multi-object tracking situations. Real-time applications are made easier by PSO's computational efficiency, which guarantees fast and rapid prediction generation. This is important in dynamic contexts where timely and accurate forecasts are necessary for making wise decisions.

PSO is attracting more academics because of its versatility and resilience, particularly for challenges in dynamic environments, it is a population-based stochastic optimization technique. Idea of PSO is the particles fly around randomly over the image searching for the best value in each round. The fittest value of the particle in its entire search at every round is called P_{Best}. The fitness function is applied to all particles, and the fitness value (best solution) is estimated and stored. The fitness value of the current optimal particle is referred to as “P_{best}”. PSO maximises the best population value gained as far as any particle in the neighbourhood, and its location is known as l_{best}.

When all of the generated populations are measured as topological neighbors by a sensitive particle, the best value is chosen among the executed population, and that sensitive best value is recognized as the best solution as g_{best}. The PSO is always attempting to change the speed of each particle towards its p_{best} and l_{best}. The speed is determined by arbitrary terms, which have arbitrarily generated quantities for the speed towards the p_{best} and l_{best} locales. The preceding stages are depicted as an Algorithm 3.

Algorithm 3: PSO Algorithm for fast visual tracking objects

```
Input: at $t_{th}$ frame of the image.
   Swarm size $(N)$, Maximum number of Iteration (M), L_rate$(L\_r)$ , acceleration coefficients $(C_1)$ ,$(C_2)$, prior width $(w)$ and height$(h)$
Initialize the image size $(n)$
For $i = 1\ to\ n$ // $i = 1, 2, 3 \ldots \ldots n$// starting with 1
        Initialize the swarm particles randomly      $SP_j = [\, X_{s1}, X_{s2}, X_{s3}, X_{s4} \ldots X_{sn}]$
// $j = 1,2,3, \ldots . N_{sp}\ j = 1,2,3, \ldots . N$, $X_1 = [x, y, w, h]$, $x$– starting x-axis, $y$-starting y-axis
      Calculate the fitness of each particle (i.e., compute the ratio classifier for all the samples
determine the $p_{best}$ and $g_{best}$ for the current round (“iteration”).
While $(t < M)$
for $j = 1\!: N_p$ //starting with_2
                    evaluate and update the velocity of the particles
                                        $V_j(t+1) = w(t+1) * V_j(t) + C_1 * rand * (P_{best} - SP_j) + C_2 * rand * (g_{best} - SP_j)$
                 $P_j(t+1) = p_j(t) + velocity\ (t+1) * lrate$ // “revise updated position”
end for // “end for 2”
    Calculate the fitness of particles $(new\_fit)$
                 revise the $P_{best}$ and $g_{best}$
        // $P_{best}$
if $(new_{fit} > P_{best}\ fit)$then
$P_{best}\ \_sol = P_j$ ;
$P_{best}\_fit\ = new\_fit$
else
                              do nothing
end if
        // $G_{best}$
if $g_{best}\_fit < max(new\_fit)$) then
$g_{best}\_Sol = P_j(index\ of\ Max(new\_fit))$
 $g_{best}\ Fit = Max(new\_fit)$
end if
end while // Repeat until the maximum num. of iterations is reached.
End for // until n frames are reached.
```

V. Experimental Results

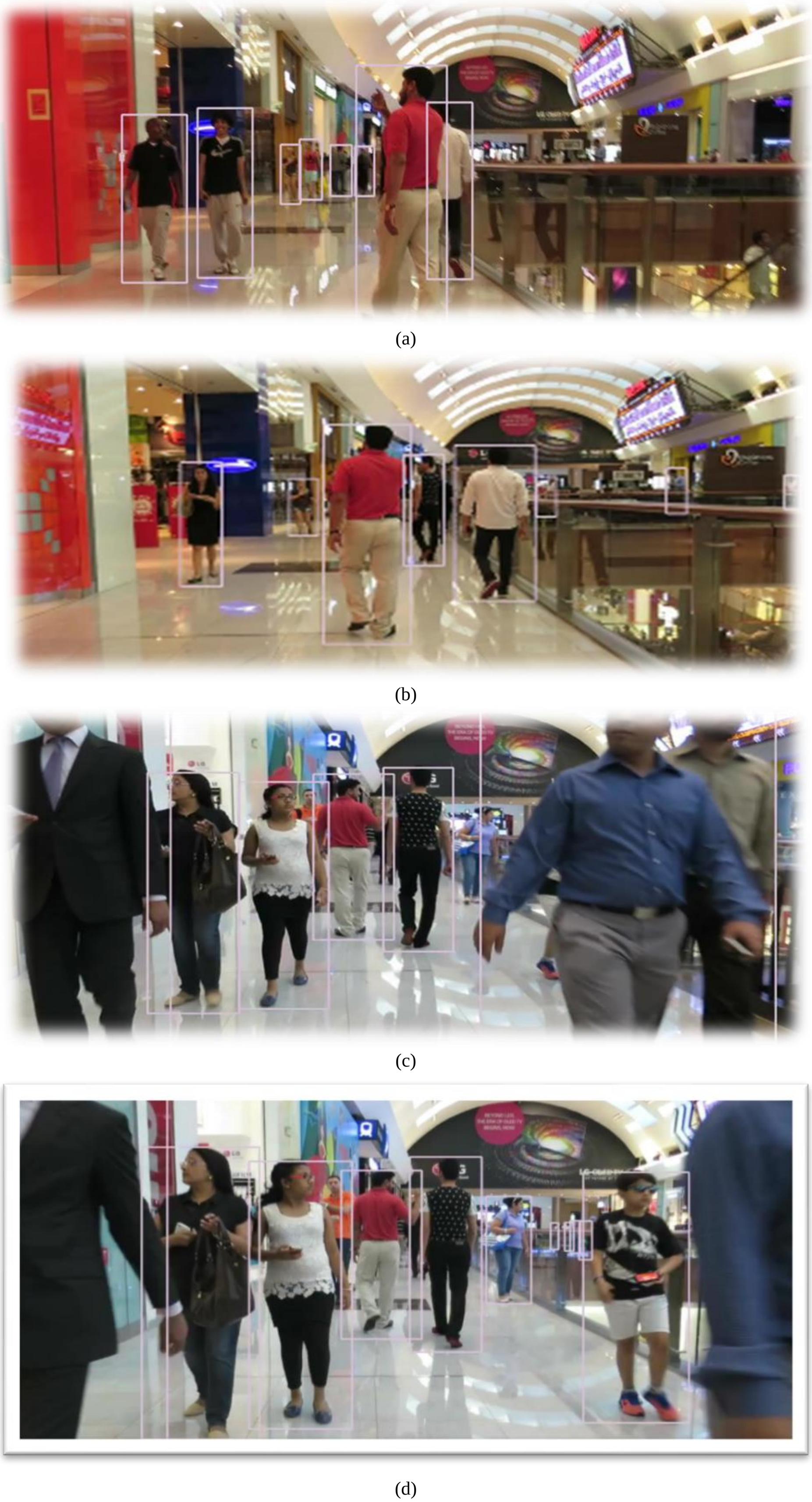
To assess the proposed methodology, MOTchallenge 2017 dataset is utilized for sample video clippings. It contains encompasses 14 series of detections, seven for training and seven for testing, thereby on a whole comprising 11245 frames. The investigations are executed in OpenCV environment, in which python is applied as the programming language.

A. Tracking Samples of the Proposed CNN-based Fast Compressive Tracking Methodology

1) Detection of people in a busy shopping mall- sample video for complete occlusion: The detection people in a shopping mall in highly occluded condition are depicted in Fig. 2(a) – (d). The video sample consists of 30 Frames Per Second (FPS), with 75 tracks and 12389 detection boxes. From Fig. 2(a) and (b) researchers can show that, the proposed methodology was able to track and detect people whose motion is unpredictable.

From Fig. 2(c) and (d) it can discovered that, the man in blue shirt has completely occluded a small boy whose shoes are only visible in the Fig. 2(c). In the frame i.e., Fig. 2(d), the boy is easily detected by the proposed methodology from which researchers can give assurance that the novel methodology is capable of detecting objects even in complete occlusion.

2) Tracking people in busy street- sample video for sudden arrival and fading of the object: Fig. 3(a), 3(b) and 3(c) depicts the example of sudden arrival and fading of the goal object. This sample video comprises of 30 FPS, with 83 tracks and 47557 detection boxes.



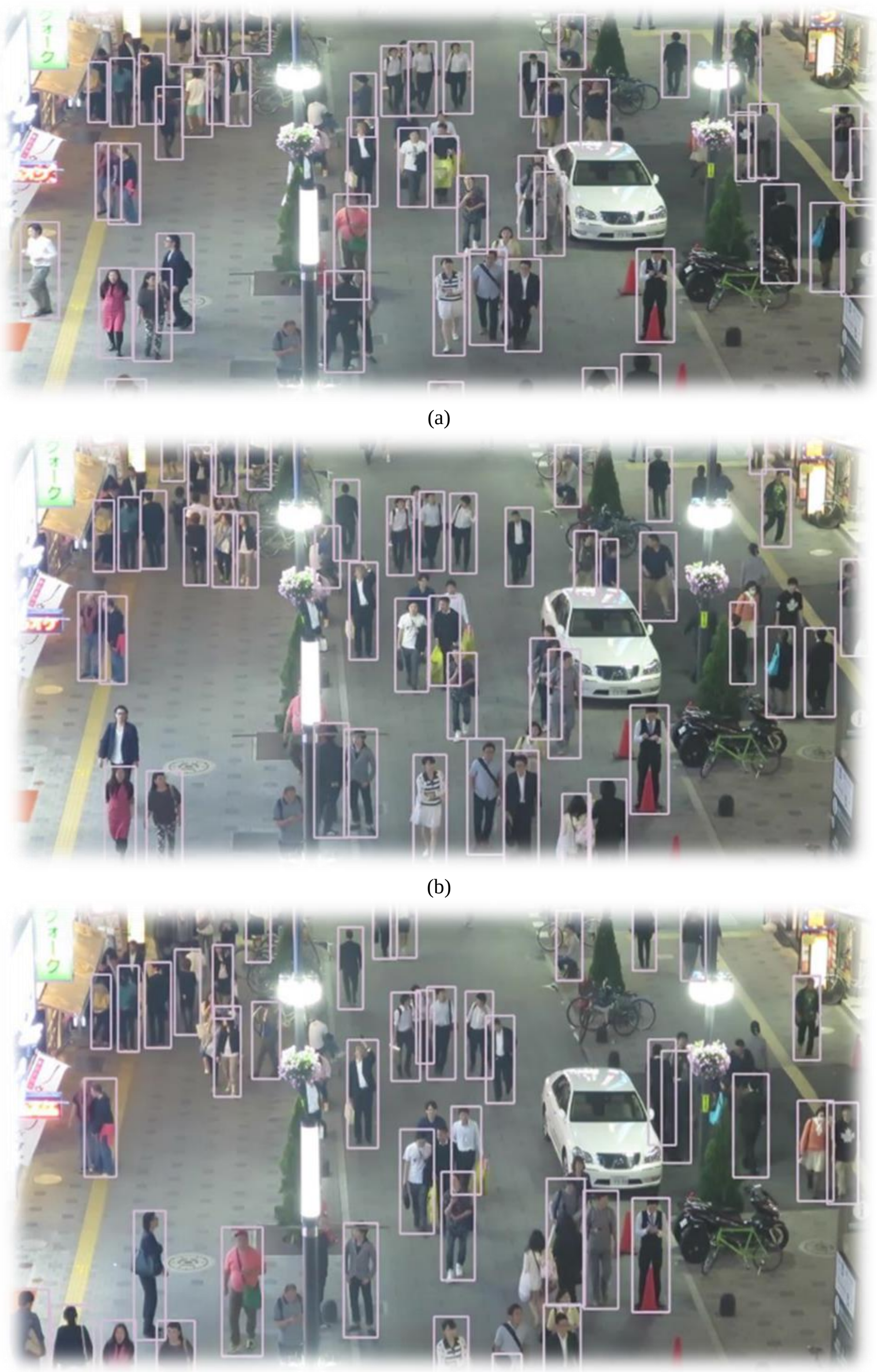
(a)

(b)

(c)

(d)

Fig. 2. (a) – (d) Detection of people in shopping malls at various frames.



(a)

(b)

(c)

Fig. 3. (a)-(c) Example of sudden appearance and disappearance of target object.

From the Fig. 3(a), it can be found that in the third street lamp, a man wearing a black hat is standing near the post. In the next Fig. 3(b), he was hidden behind the post. After sometimes, he is again appearing and tracked in the Fig. 3(c). Thus the proposed methodology was able to track objects with sudden appearance and re-appearance behavior.

Researchers are evaluating using MOT as a performance metric suite, which comprises MOT_A (Multi-Object Tracking Accuracy) and MOT_P (Multi-Object Tracking Precision) as the chief metrics The MOT_A collects “three error causes: False Positives (FP), False Negatives (FN) and Identity Switches (IDs)”. “*Gt* is the sum of the goals in each edge. Researchers also investigate about the sum of courses of Ground_Truth (GT), Mostly_Tracked (MT), Mostly_Lost (ML), Fragment (FM), Partially_Tracked trajectories (PT), and Multiple Object Tracking Accuracy with log 10 (*IDs*) (MOTAL), False_Alarms per Frame (FAF). Researchers also acknowledge the assessment metrics Recall (Rcll) and Precision (Prcn)”. Each technique's rate is restrained in frames per second (FPS). The subsequent table compares several object followers in the MOT2017 encounter dataset to the planned scheme.

The above Table I make a comparison between the existing multi-objects tracking to that of the proposed methodology. The comparison was made taking two datasets which was used as sample video sequences for the experimental purpose namely, shopping mall dataset and busy street dataset. From the table it can be concluded that, the proposed methodology showed better performance against MOT2017 dataset state-of-arts methods when compared to that of existing object tracking methodologies.

The graphical depiction of the comparison between the proposed CNN-FCT and M-FCT is shown in the following graphs. Fig. 4 compares the Accuracy, Robustness and efficiency of the proposed methodology with that of M-FCT. It was found that, CNN-FCT showed enhanced performance measures when compared to that of M-FCT.

Fig. 5 depicts the MOT accuracy between CNN-FCT and M-FCT, whereas Fig. 6 depicts the Most Tracked and Most Least measures of the proposed and existing methodologies. As the proposed CNN-FCT tracks object even if the object is fully occluded or appears and re-appears randomly, the MOT Accuracy is higher than that of M-FCT. Even though M-FCT is better when compared to conventional object tracking methods, its track loss rate is higher when compared to that of CNN-FCT. Hence from the comparisons, it can be evident that the proposed methodology performed well against all recently proposed multi-object tracking algorithms, especially in case of complete occlusion and object appearance and disappearance scenario.

B. *Performance Metrics Evaluation*

Quantitative indicators that evaluate accuracy and effectiveness are referred to as metrics of performance. These measurements are employed to assess and contrast various tracking approaches or algorithms. Here are a few performance measures that are frequently employed in object tracking such as recall, mAP, F1-score, Peak Signal to Noise Ratio (PSNR) and Root Mean Square Error (RMSE).

1) Recall: The proportion of real optimistic outcomes that a model correctly organizes as being optimistic is known as recall. The proportion of true positives to the entire of true positives and false negatives is secondhand to calculate it. It is given in Eq. (2) below:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{2}$$

TABLE. I MOT2017 CHALLENGE DATASET MULTI-OBJECT TRACKING VERSUS THE PROPOSED METHODOLOGY

Datasets	Methods	MOT-A	MOT_P	MTrack	MLost	FP(+ve)	FN (-ve)	IDs (/Rcll)	FM(/Rcll)	Speed
Mall dataset	ACF [30]	33.7	76.5	7.2%	54.2%	5,804	112,587	2,418 (63.2)	2,252 (58.9)	1.3
	ZF [30]	33.2	75.5	7.8%	54.4%	6,837	114,322	642 (17.2)	731 (19.6)	0.3
	CNN-FCT (proposed)	**40.9**	**88.8**	**9.8%**	**34.0%**	**3,356**	**83,108**	**567 (11.78)**	**670 (18.4)**	**12.5**
Busy Street dataset	ACF [30]	29.7	75.2	5.3%	47.7%	17,426	107,552	3,108 (75.8)	4,483 (109.3)	0.2
	ZF [30]	26.2	76.3	4.1%	67.5%	3,689	130,54	365 (12.9)	638 (22.5)	22.2
	CNN-FCT (proposed)	**38.4**	**90.34**	**9.7%**	**36.0%**	**3,378**	**85,108**	**572 (12.78)**	**620(18.4)**	**13.6**

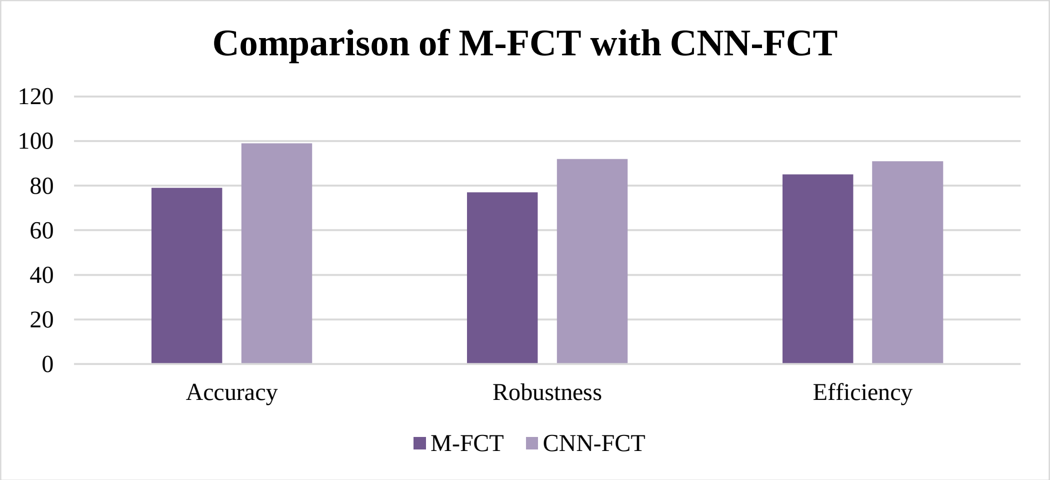


Fig. 4. Performance metrics comparison graph.

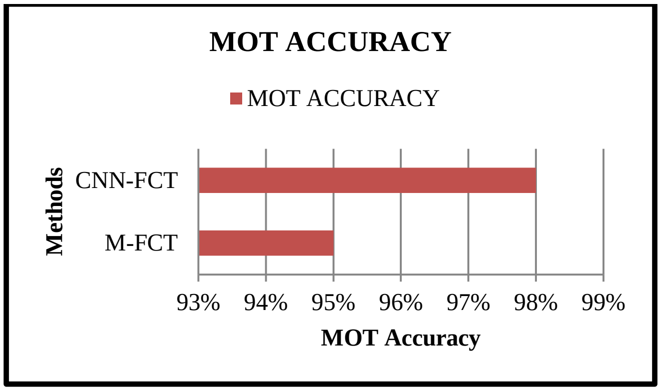


Fig. 5. Mot accuracy.

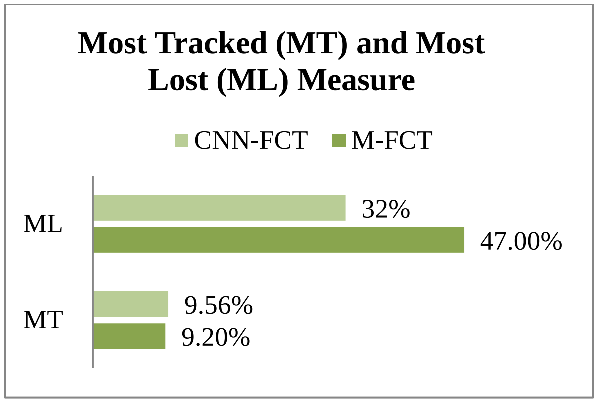


Fig. 6. ML and MT measures.

2) mAP: The mean over every category is then calculated by mAP after calculating the standard deviation of precision for every class. Its equation is given in Eq. (3) below,

$$mAP = \sum_{p=1}^{P} \frac{Average\ Q(p)}{P} \quad (3)$$

3) F1 Score: The F1 score is an individual statistic that syndicates precision and recall. It is the vocal average of these two metrics. It is frequently employed in binary categorization jobs where the proportion of both positive and negative instances is unbalanced. Its equation is given in Eq. (4) below,

$$\text{F1-Score} = \frac{\text{2 x Precision x Recall}}{\text{Precision+Recall}} \quad (4)$$

4) Root Mean Square Error (RMSE): A standard metric for assessing the effectiveness of models of regression is RMSE. By taking into account the squared variations, it calculates the average variance among the expected and actual outcomes. When greater errors are more important, RMSE is especially helpful since it draws attention to more significant discrepancies. It is given in Eq. (5) below,

$$RMSE = \sqrt{\sum_{j=1}^{M} \frac{\|x(j)-\hat{x}(j)\|}{N}} \quad (5)$$

Here, j is represented as the variable; M is represented as the non-missing data points; $x(j)$ is represented as the actual observation time series; $\hat{x}(j)$ is represented as the estimated time series.

5) Peak Signal to Noise Ratio (PSNR): A commonly used statistic for assessing the effectiveness of video or image compression methods is PSNR. It calculates the difference among the highest possible signal strength and the background noise power. It is given in Eq. (6) below,

$$PSNR = \frac{10\ log_{10}\ (peak\ value)^2}{MSE} \quad (6)$$

TABLE. II COMPARISON TABLE OF RECALL, MAP, F1 SCORE

Method	Recall	Map	F1 score
Clustering Method [23]	71	53	61
YOLOV4 DNN Model [24]	93	96	93
YOLOV3 Model [25]	41	46	53
Proposed CNN-FCT Model	95	96.9	95.3

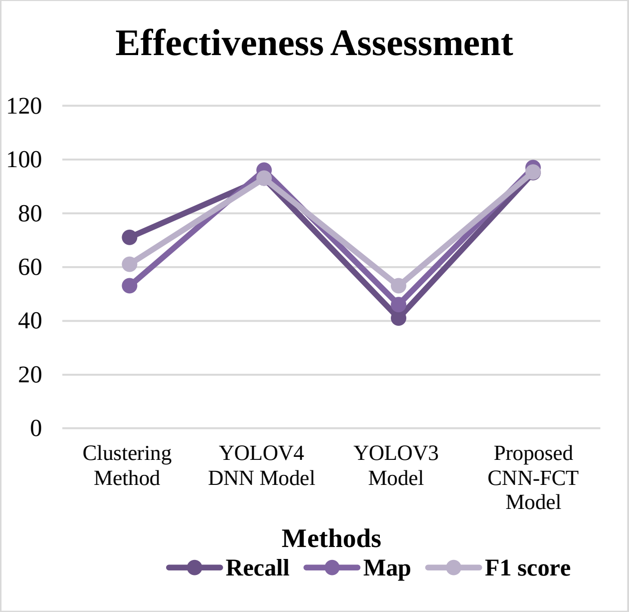


Fig. 7. Effectiveness assessment of Recall, mAP and F1-Score.

Table II makes the assessment among the existing methods recall, mAP and F1 score with the proposed CNN-FCT method. The proposed model produces greater recall, mAP and F1 score points and Fig. 7 depicts the comparison graph of the existing methods recall, mAP and F1 score with the proposed CNN-FCT method.

Other metrics like RMSE and PSNR also takes place and the below Table III shows the assessment table of it and Fig. 8 shows the comparison graph of the existing methods RMSE and PSNR with the proposed CNN-FCT method.

TABLE. III COMPARISON TABLE OF RMSE AND PSNR

Method	RMSE	PSNR
CNN Method [26]	24.09	30.38
OTSU Algorithm [27]	43.18	29.16
Proposed CNN-FCT Model	18.25	35.72

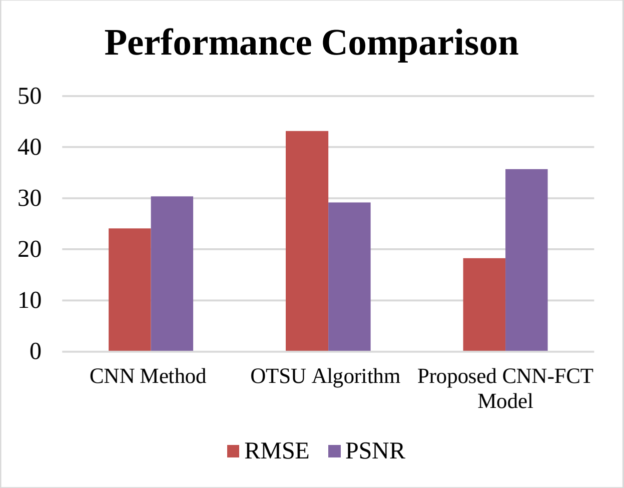


Fig. 8. Comparison graph of RMSE and PSNR.

The phrase fitness improvement over iteration in this context refers to the detection method's use of an optimization iteration procedure that enhances the precision and effectiveness of various item monitoring. A Convolutional Neural Network (CNN) model, which serves as an extractor of features for representing the objects in the video frames, is at the core of the approach. The CNN algorithm extracts the object's features after estimating the object's location in the initial frame during the tracking procedure. The predicted location of the first frame might not be completely precise, leading to a less-than-ideal initial solution. The Kalman Filter and PSO algorithm are used to modify the object's projected position at every iteration, and the fitness metric is calculated as well. This fitness statistic measures how closely the predicted position of the object matches the actual or intended position. Fig. 9 shows the fitness improvement graph.

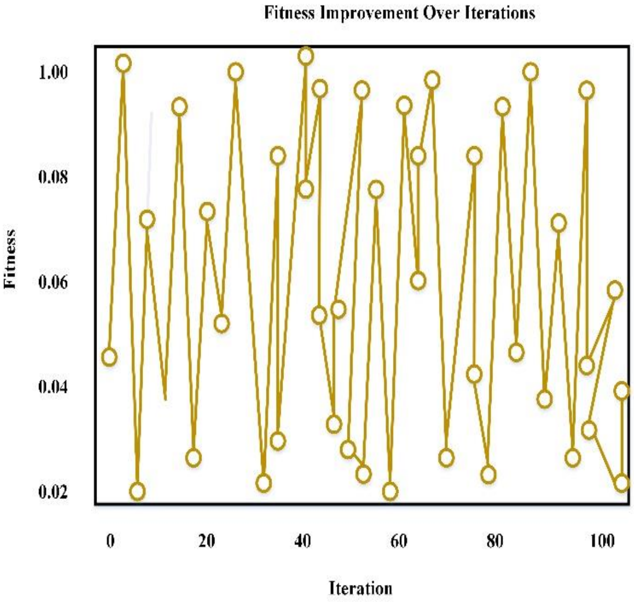


Fig. 9. Fitness improvement over iteration graph.

The graph contrasting the efficacy of a model with and without Accuracy, Recall, mAP (mean Average Precision), and F1 score optimization offers insightful information about the importance of focused improvement in particular machine learning tasks. The algorithm undergoes training in the with optimization case primarily deals with the goal of enhancing The key performance indicators for applications like object identification and classification are accuracy, recall, mAP, and F1 score. The graph shows a stronger upward trend as a result of the model's constant fine-tuning to maximize these metrics. When there is no optimization, the algorithm is trained without giving these metrics any special consideration, which results in more or less noticeable improvements in Accuracy, Recall, mAP, and F1 score. The graphs of accuracy, recall, mAP, and F1 score with and without optimization are displayed in Fig. 10.

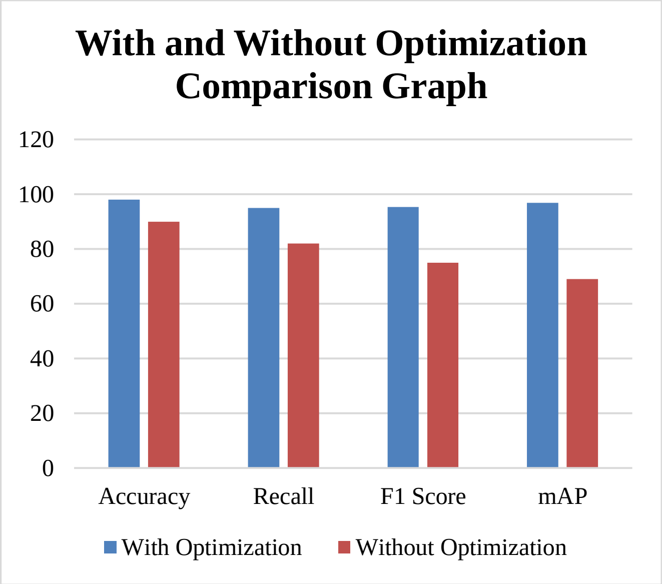


Fig. 10. With and without optimization comparison graph of accuracy, Recall, mAP, F1 score.

C. Discussion

In order to improve object tracking, the research article suggests a Fast Compressive Tracking approach based on Convolutional Neural Network (CNN) models. The CNN model is fed two sets of the dimensionally decreased object characteristics using this method. The CNN algorithm analyses the input frames via consecutive layers of the CNN model, removing the salient characteristics from the input frames. The Particle Swarm Optimization (PSO) technique is used by the CNN model's output coating to record the positions of the monitored targeted image features. The intended image is then classified using a Support Vector Machine (SVM) classifier based on the monitored positions. Here, the planned CNN based fast visual tracking architecture is depicted in Fig. 1. The Multi Object FCT algorithm is given in Algorithm 1. CNN based Tracking algorithm is given in Algorithm 2. Then, PSO algorithm for fast visual tracking object is given in Algorithm 3. In the results section, the detection people in a shopping mall in highly occluded condition are depicted in Fig. 2(a) – (d). The video sample consists of 30 Frames Per Second (FPS), with 75 tracks and 12389 detection boxes. From Fig. 2(a) and (b) researchers can show that, the proposed methodology was able to track and detect people whose motion is unpredictable. Then, Fig. 2(c) and (d) it can discover that, the man in blue shirt has completely occluded a small boy whose shoes are only visible in the fig. 2(c). Fig. 2(d), shows the boy is easily detected by the proposed methodology from which researchers can give assurance that the novel methodology is capable of detecting objects even in complete occlusion. Fig. 3(a), 3(b) and 3(c) depicts the example of sudden appearance and disappearance of the target object. This sample video comprises of 30 FPS, with 83 tracks and 47557 detection boxes. After that, Table I makes a comparison between the existing multi-objects tracking to that of the proposed methodology. The comparison was made taking 2 datasets which was used as sample video sequences for the experimental purpose namely, shopping mall dataset and busy street dataset. Fig. 5 depicts the MOT accuracy between CNN-FCT and M-FCT, whereas Fig. 6 depicts the Most Tracked and Most Least measures of the proposed and existing methodologies. Table II makes the comparison between the existing methods recall, mAP and F1 score with the proposed CNN-FCT method. The proposed model produces greater recall, mAP and F1 score points and Fig. 7 depicts the comparison graph of the existing methods recall, mAP and F1 score with the proposed CNN-FCT method. Table III displays the assessment table of it and Fig. 8 displays the assessment graph of the existing methods RMSE and PSNR with the planned CNN-FCT method [26] [27]. Fig. 9 shows the fitness improvement graph and it is used to determine a performance achievement's advantages or flaws. Fig. 10 shows the with and without optimization graph of Accuracy, Recall, mAP and F1 score because it is important to remember that these conclusions are hypothetical considering the limited facts at hand [23] [24] [25].

Compared to previous multi-object identification techniques, the CNN-FCT Methodology works well because it addresses these drawbacks. Real-time tracking may be difficult for traditional methods, particularly in situations where there is occlusion or abrupt object movements. In order to overcome these drawbacks, the CNN-FCT Methodology incorporates Kalman Filtering, FCT, CNN, and PSO. This results in improved accuracy, flexibility, and resilience while dealing with changing environmental circumstances. As a promising development in the realm of multi-object identification, the methodology stands out for its capacity to tackle a variety of obstacles.

The proposed CNN-FCT technology in addition to displaying the experimental results and performance metrics. The theoretical advances and computing efficiency attained by the optimization processes have immediate real-world applications in object tracking scenarios such as surveillance systems, traffic monitoring, and human activity analysis. Because of the durability exhibited in dealing with hard conditions such as occlusion and the unexpected appearance/disappearance of objects, this technology is particularly helpful for practical application in dynamic contexts. Enhanced measures such as accuracy, recall, mAP, and F1 score directly contribute to improving the dependability and efficacy of object tracking systems across multiple domains. The proposed methodology shows itself as a viable tool for real-world applications by stressing these practical consequences, addressing the demand for improved and efficient object tracking systems.

VI. Conclusion and Future Work

The combined strategy of using Particle Swarm Optimization (PSO), Kalman Filtering, Convolutional Neural Networks (CNNs), and Compressive Sensing has shown to be very successful in addressing the difficulties involved in multi-object tracking and identification in dynamic environments. By using CNNs for accurate object recognition, compressive sensing for well-tuned representations, Kalman Filtering for flexibility, and PSO for the best forecasting accuracy, this synergistic system performs better than individual techniques. The joint application of these methods showed excellent tracking performance under difficult conditions, such as complex motion patterns, occlusions, and visual discrepancies. This suggested methodology has notable practical benefits. With a tracking accuracy of 98%, the system demonstrates its potential for practical use, especially in fields that demand advanced multi-object recognition and tracking. In addition to advancing computer vision, the study establishes a strong basis for further advancements in multi-object tracking systems. It is important to recognize some restrictions, though. Real-time monitoring in resource-constrained contexts requires further optimization of the computational efficiency of the program. Investigating transfer learning and domain adaptation can improve the system's generalization in a variety of tracking environments. Three-dimensional object tracking would benefit from the addition of depth information, and more research is necessary to determine how robust the method is against inclement weather and abrupt changes in illumination. Future work can concentrate on optimizing the methodology for real-time application, increasing scalability to support a larger number of objects and a wider range of surroundings, and combining it with creative algorithms. The above efforts would certainly improve the capabilities of the methodology and increase its application across many areas. To sum up, our research offers significant understandings of the theoretical and practical aspects of multi-object tracking systems, opening new avenues for further investigation and development in computer vision. The suggested CNN-FCT methodology performs well in a variety of settings, but it is important to recognize its possible advantages as well as its limitations with regard to different kinds of data. To further improve adaptability and meet the particular obstacles presented by various circumstances, future study could investigate algorithm adaptations or parameter tailoring for certain data kinds.

References

[1] F. Schilling, F. Schiano, and D. Floreano, "Vision-Based Drone Flocking in Outdoor Environments," IEEE Robot. Autom. Lett., vol. 6, no. 2, pp. 2954–2961, Apr. 2021, doi: 10.1109/LRA.2021.3062298.

[2] A. A. Rafique, A. Jalal, and A. Ahmed, "Scene Understanding and Recognition: Statistical Segmented Model using Geometrical Features and Gaussian Naïve Bayes," in 2019 International Conference on

Applied and Engineering Mathematics (ICAEM), Taxila, Pakistan: IEEE, Aug. 2019, pp. 225–230. doi: 10.1109/ICAEM.2019.8853721.

[3] W. Ren, X. Wang, J. Tian, Y. Tang, and A. B. Chan, “Tracking-by-Counting: Using Network Flows on Crowd Density Maps for Tracking Multiple Targets,” IEEE Trans. Image Process., vol. 30, pp. 1439–1452, 2021, doi: 10.1109/TIP.2020.3044219.

[4] A. Farid, F. Hussain, K. Khan, M. Shahzad, U. Khan, and Z. Mahmood, “A Fast and Accurate Real-Time Vehicle Detection Method Using Deep Learning for Unconstrained Environments,” Appl. Sci., vol. 13, no. 5, p. 3059, Feb. 2023, doi: 10.3390/app13053059.

[5] J. Zhou, S. Zeng, and B. Zhang, “Linear Representation-Based Methods for Image Classification: A Survey,” IEEE Access, vol. 8, pp. 216645–216670, 2020, doi: 10.1109/ACCESS.2020.3041154.

[6] N. Sindhwani, R. Anand, M. S., R. Shukla, M. Yadav, and V. Yadav, “Performance Analysis of Deep Neural Networks Using Computer Vision,” EAI Endorsed Trans. Ind. Netw. Intell. Syst., vol. 8, no. 29, p. 171318, Nov. 2021, doi: 10.4108/eai.13-10-2021.171318.

[7] W.-G. Li and H. Wan, “An improved spatio-temporal context tracking algorithm based on scale correlation filter,” Adv. Mech. Eng., vol. 11, no. 2, p. 168781401982590, Feb. 2019, doi: 10.1177/1687814019825903.

[8] C. R. Bern, K. Walton-Day, and D. L. Naftz, “Improved enrichment factor calculations through principal component analysis: Examples from soils near breccia pipe uranium mines, Arizona, USA,” Environ. Pollut., vol. 248, pp. 90–100, May 2019, doi: 10.1016/j.envpol.2019.01.122.

[9] H. D. Najeeb and R. F. Ghani, “A Survey on Object Detection and Tracking in Soccer Videos,” Muthanna J. Pure Sci., vol. 8, no. 1, pp. 1–13, Jan. 2021, doi: 10.52113/2/08.01.2021/1-13.

[10] Assistant Professor, Department of Computer Science and Engineering, IFET College of Engineering, Villupuram, India and Dr. D. S. David*, “An Intellectual Individual Performance Abnormality Discovery System i n Civic Surroundings,” Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 5, pp. 2196–2206, Mar. 2020, doi: 10.35940/ijitee.E2133.039520.

[11] M. M. Moussa and R. Shoitan, “Object-based video synopsis approach using particle swarm optimization,” Signal Image Video Process., vol. 15, no. 4, pp. 761–768, Jun. 2021, doi: 10.1007/s11760-020-01794-1.

[12] N. Nedjah, A. V. Cardoso, Y. M. Tavares, L. D. M. Mourelle, B. B. Gupta, and V. Arya, “Co-Design Dedicated System for Efficient Object Tracking Using Swarm Intelligence-Oriented Search Strategies,” Sensors, vol. 23, no. 13, p. 5881, Jun. 2023, doi: 10.3390/s23135881.

[13] X. Chen, F. Li, and X. Liu, “Efficient and Robust Distributed Digital Codec Framework for Jointly Sparse Correlated Signals,” IEEE Access, vol. 7, pp. 77374–77386, 2019, doi: 10.1109/ACCESS.2019.2920982.

[14] D. T. Nguyen, T. N. Nguyen, H. Kim, and H.-J. Lee, “A High-Throughput and Power-Efficient FPGA Implementation of YOLO CNN for Object Detection,” IEEE Trans. Very Large Scale Integr. VLSI Syst., vol. 27, no. 8, pp. 1861–1873, Aug. 2019, doi: 10.1109/TVLSI.2019.2905242.

[15] D. Li, Y. Yu, and X. Chen, “Object tracking framework with Siamese network and re-detection mechanism,” EURASIP J. Wirel. Commun. Netw., vol. 2019, no. 1, p. 261, Dec. 2019, doi: 10.1186/s13638-0191579-x.

[16] S. Kaderali, “Detection and Characterisation of Unknown Spacecraft Maneuvers,” 2021.

[17] I. Ullah, X. Su, J. Zhu, X. Zhang, D. Choi, and Z. Hou, “Evaluation of Localization by Extended Kalman Filter, Unscented Kalman Filter, and Particle Filter-Based Techniques,” Wirel. Commun. Mob. Comput., vol. 2020, pp. 1–15, Oct. 2020, doi: 10.1155/2020/8898672.

[18] N. Mahmoudi, S. M. Ahadi, and M. Rahmati, “Multi-target tracking using CNN-based features: CNNMTT,” Multimed. Tools Appl., vol. 78, no. 6, pp. 7077–7096, Mar. 2019, doi: 10.1007/s11042-018-6467-6.

[19] Y. Liu et al., “FedVision: An Online Visual Object Detection Platform Powered by Federated Learning,” Proc. AAAI Conf. Artif. Intell., vol. 34, no. 08, pp. 13172–13179, Apr. 2020, doi: 10.1609/aaai.v34i08.7021.

[20] C. Li and B. Yang, “Adaptive Weighted CNN Features Integration for Correlation Filter Tracking,” IEEE Access, vol. 7, pp. 76416–76427, 2019, doi: 10.1109/ACCESS.2019.2922494.

[21] P. Dendorfer et al., “MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking,” Int. J. Comput. Vis., vol. 129, no. 4, pp. 845–881, Apr. 2021, doi: 10.1007/s11263-020-01393-0.

[22] “Papers with Code - MOT17 Dataset.” https://paperswithcode.com/dataset/mot17 (accessed Aug. 14, 2023).

[23] L. Zhao and S. Li, “Object Detection Algorithm Based on Improved YOLOv3,” Electronics, vol. 9, no. 3, p. 537, Mar. 2020, doi: 10.3390/electronics9030537.

[24] A. M. Roy, R. Bose, and J. Bhaduri, “A fast accurate fine-grain object detection model based on YOLOv4 deep neural network.” arXiv, Oct. 30, 2021. Accessed: Jul. 18, 2023. [Online]. Available: http://arxiv.org/abs/2111.00298

[25] U. Nepal and H. Eslamiat, “Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs,” Sensors, vol. 22, no. 2, p. 464, Jan. 2022, doi: 10.3390/s22020464.

[26] F. Klepel and R. Goebel, “Learning Equivariant Object Recognition and its Reverse Application to Imagery,” Neuroscience, preprint, May 2023. doi: 10.1101/2023.05.20.541553.

[27] M. Wang, M. Lv, H. Liu, and Q. Li, “Mid-Infrared Sheep Segmentation in Highland Pastures Using Multi-Level Region Fusion OTSU Algorithm,” Agriculture, vol. 13, no. 7, p. 1281, Jun. 2023, doi: 10.3390/agriculture13071281.

Comparative Analysis of Weighted Ensemble and Majority Voting Algorithms for Intrusion Detection in OpenStack Cloud Environments

Pravin Patil, Geetanjali Kale, Nidhi Bivalkar, Agneya Kolhatkar
Department of Computer Engineering, Pune Institute of Computer Technology, Pune, Maharashtra, India

Abstract—**In the ever-evolving landscape of cybersecurity, the detection of malicious activities within cloud environments remains a critical challenge. This research aims to compare the effectiveness of two ensemble algorithms, the weighted ensemble algorithm and the majority voting algorithm, in the context of intrusion detection within an OpenStack cloud environment. To conduct this study, a dataset was generated using a network of 10 virtual machines, simulating the complex dynamics of a real cloud infrastructure. Various attack scenarios were simulated, and system metrics including CPU usage, memory utilization, and network traffic were monitored and logged. The weighted ensemble algorithm combines the predictions of multiple individual models with varying weights, while the majority voting algorithm aggregates predictions from multiple models. Through a rigorous experimental setup, these algorithms were applied to the generated dataset, and their performance was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These findings provide valuable insights into the strengths and weaknesses of ensemble algorithms for intrusion detection in cloud environments. It highlights the importance of selecting appropriate algorithms based on specific security requirements and threat profiles. Different attack scenarios may require different algorithmic approaches to achieve optimal results. Overall, this study contributes to the understanding of ensemble techniques in cloud security and offers a foundation for further research in optimizing intrusion detection strategies within dynamic and complex cloud environments. By identifying the strengths and weaknesses of different ensemble algorithms, cybersecurity professionals can make informed decisions in selecting the most suitable approach to enhance the security of cloud environments.**

Keywords*—*Intrusion detection; ensemble algorithms; cloud security; openstack; weighted ensemble; majority voting

I. Introduction

The advancement of computing has ushered in transformative technologies, with cloud environments being at the forefront. These digital ecosystems offer unparalleled convenience, scalability, and connectivity, fundamentally reshaping the storage and processing of data. However, along with these advantages comes the pressing challenge of securing data within interconnected cloud systems. These systems, while efficient, create pathways for a diverse range of cyber threats that require robust and adaptable intrusion detection mechanisms. In the pursuit of fortifying cloud security, traditional intrusion detection approaches have played a crucial role [1], [3]. Nevertheless, these methodologies face limitations as attackers continuously evolve their tactics. Rule-based systems prescribe inflexible attack patterns, signature detection relies on pre-identified attack signatures, and anomaly detection, while effective against new attacks, often yields high false positive rates. To overcome these limitations, the integration of ensemble techniques into intrusion detection systems (IDS) has emerged as a promising strategy. Ensembles amalgamate the insights of multiple models to enhance accuracy and robustness, enabling systems to adapt to evolving attack strategies. Within the realm of ensembles, two methods stand out prominently: the weighted ensemble and the majority voting algorithms.

This research embarks on a comprehensive exploration of these two algorithms within the dynamic framework of OpenStack cloud environments. OpenStack, renowned as an open-source cloud platform, provides an intricate architecture and susceptibility to real-world cyber threats, making it an ideal evaluation ground for ensemble techniques. At the heart of our investigation lies the question of which ensemble algorithm, between the weighted ensemble and majority voting, demonstrates superior performance in the field of intrusion detection within OpenStack environments. To address this question, our methodology involves meticulously simulating a wide range of attack scenarios within the OpenStack ecosystem. By creating a synthetic environment consisting of virtual machines that mimic the complexities of cloud ecosystems, we subject our algorithms to various cyber-attacks. This deliberate diversification encompasses different tactics and vectors, resulting in a comprehensive evaluation of the algorithms' resilience and adaptability. Furthermore, during these simulations, we diligently record and analyze intricate system metrics. This detailed dataset sheds light on how the algorithms behave under dynamic attack conditions, enhancing our understanding of their effectiveness and response. In essence, the main contribution of this research lies in the empirical evaluation of the weighted ensemble and majority voting algorithms. Through rigorous experimentation and thorough analysis of the results, we uncover valuable insights into their operational dynamics, strengths, and limitations. Additionally, our findings have practical implications for the real-world deployment of these algorithms within intrusion detection systems. Fig. 1 gives an overall research workflow.

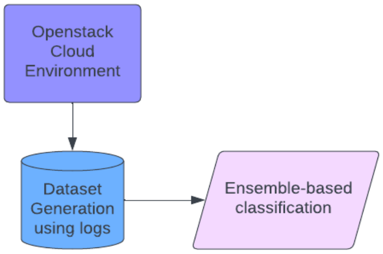

Openstack Cloud Environment
Dataset Generation using logs
Ensemble-based classification

Fig. 1. Workflow.

This research paper unfolds in several sequential sections following the introduction. The Related Work section delves into existing intrusion detection systems, thereby establishing the groundwork for ensemble techniques. Following this, the Data Collection section provides an in-depth account of the creation of a synthetic dataset, designed to simulate a wide range of cyber-attacks within an OpenStack environment. Subsequently, in the Classification Algorithms section, ten different supervised learning models are introduced, serving as the foundation for subsequent evaluations of ensemble techniques. The obtained results yield a comprehensive performance analysis, comparing various metrics such as accuracy, precision, recall, and F1 score and further implications of those results are discussed. Finally, the Conclusion synthesizes the findings, highlighting the strengths of the research and proposing potential avenues for future investigations.

II. Related Work

The continuous pursuit of enhancing intrusion detection systems has propelled researchers to explore a diverse range of methodologies. Traditional approaches encompass rule-based systems, signature detection, and anomaly detection. Rule-based systems prescribe static attack patterns, yet struggle to accommodate the dynamic nature of evolving attack strategies.

Signature detection relies on predefined attack signatures, rendering it ineffective against novel attacks that evade established patterns [1], [4]. Wie et al. and W. Hu et al. used the AdaBoost Classifier for a Network IDS [5], [6]. A. Rai et al. designed optimized IDS using deep neural networks and the GradientBoost Classifier [7].

Ensemble methods leverage the collective insights of multiple models to enhance accuracy and resilience, enabling systems to adapt to emerging attack tactics. Notably, the weighted ensemble and majority voting algorithms have emerged as formidable contenders within the realm of ensemble-based intrusion detection. The weighted ensemble algorithm hinges on the principle of model weighting, dynamically assigning importance to individual model predictions [8]. This adaptability empowers the algorithm to excel across diverse attack scenarios, optimally adjusting the influence of each model based on its performance characteristics.

Conversely, the majority voting algorithm capitalizes on the synthesis of predictions from multiple models [9]. By establishing a consensus among models, this approach fosters robustness, mitigating the impact of errors arising from individual models. Several studies have explored the applications of ensemble techniques for intrusion detection [8], [2] [10], [9], [11], [12].

There has also been significant research on the applications of these classifiers on multi-tenant systems [8], [9]. While these previous studies have enriched the discourse on ensemble-based intrusion detection, the comparative assessment of the weighted ensemble and majority voting algorithms remains a relatively unexplored area, particularly within the dynamic context of OpenStack cloud environments [4]. It is within this realm that our research finds its foundation, systematically evaluating the performance of these algorithms in a relevant cloud security landscape. Previous studies have shed light on the potential of ensemble techniques. However, these studies often lack a comprehensive analysis of the strengths and weaknesses of the algorithms. The specific limitations of each method, particularly within the OpenStack environment, have not been thoroughly addressed. To bridge this gap, our research undertakes an extensive comparative analysis of the weighted ensemble and majority voting algorithms. By subjecting these algorithms to attack scenarios within OpenStack cloud environments, we aim to discern their nuanced responses and understand their operational dynamics in the face of complex threats. Through this exploration, our study strives to offer practical insights into the adaptability and effectiveness of these algorithms, enabling informed decision-making for intrusion detection in multi-tenant distributed systems.

III. Data Collection

The research methodology encompasses a seamless continuum from data collection to model training. System metrics are meticulously collected from a simulated OpenStack cloud, replicating real-world dynamics during attacks. The raw data undergoes thorough preprocessing, including cleaning and feature engineering, transforming the metrics into meaningful attributes for insightful analysis. The refined dataset is used to train weighted ensemble and majority voting models. The models undergo iterative adjustments and fine-tuning to optimize their intrusion detection capabilities. This integrated process establishes a robust evaluation framework for ensemble algorithms within the intricate context of OpenStack cloud scenarios.

A. *Dataset Generation*

In our pursuit of conducting a comprehensive evaluation, we systematically undertook the task of creating a virtual cloud environment utilizing the OpenStack framework. This task utilized the computational resources of two laptops, one with 8GB of RAM and the other with a substantial 32GB. Within this setup, we deployed ten strategically distributed virtual machines (VMs) across the laptops, each serving as a control node or a compute node. These roles mirrored the dynamics of a real-world cloud ecosystem and facilitated a highly realistic evaluation [13]. To further enhance the realism of our evaluation, we simulated cyber-attacks on this virtual cloud environment, with a focus on the control node, a vital component of the cloud infrastructure. The attacks encompassed a wide range of threats, including Cross-Site Request Forgery (CSRF), XML External Entity (XXE) Injection, Brute Force, Cross-Site Scripting (XSS), Open Redirect, Directory Traversal, SQL Injection, Command Injection, and Remote Code Execution. Each attack category had the potential to impact various critical parameters within the cloud environment. Each attack category had distinct implications for the cloud environment, leading to specific performance metrics. Each attack category had the potential to influence these system performance metrics, resulting in nuanced impacts on the operational landscape of the cloud environment.

1) CSRF could compromise data integrity and disrupt user interactions, potentially affecting transaction success rates and request latencies [14].

2) XEE could impact system behaviour and data confidentiality, potentially influencing system response times and memory utilization [15].

3) Brute Force attacks on authentication mechanisms could have system-wide consequences, impacting authentication failure rates and overall system availability.

4) XSS could undermine user interactions and data integrity, posing threats to user session durations and engagement metrics.

5) Open Redirect vulnerabilities could impact user navigation experiences, potentially affecting click-through rates and user satisfaction levels.

6) Directory Traversal exploits could influence file access rates and disk I/O operations.

7) SQL Injection could jeopardize data confidentiality and system availability, influencing query execution times and database throughput.

8) Command Injection could manipulate system commands, potentially affecting CPU usage and system response times.

9) Remote Code Execution posed severe risks to system integrity and availability, potentially impacting memory usage and network traffic rates.

The execution of these attacks was randomized to ensure a diverse range of threat scenarios. To capture the dynamics of the cloud environment during attacks, we utilized the Netdata REST API service to collect real-time system metrics. These metrics were meticulously organised into a structured CSV format for subsequent analysis. The extent of our evaluation is evident in the execution of a total of 10,000 attack instances, showcasing the rigorous and dedicated nature of our study. This comprehensive exploration serves as a robust foundation for analyzing the performance of ensemble algorithms in real-world scenarios.

The distribution of attack and non-attack instances is clearly shown in Table I.

The dataset utilized in this study comprises a diverse range of system metrics and attributes collected from a simulated OpenStack cloud environment. With a total of 63 distinct parameters, each column represents a specific system parameter or feature [16]. These parameters encompass various measurements related to CPU utilization, memory consumption, disk activity, network behaviour, process behaviour, firewall activity, and more. Each row in the dataset corresponds to a specific time instance during simulated attack scenarios. A value of 1 or 0 is assigned to categorise the instances, where 1 denotes an attack instance and 0 represents a non-attack instance. For attack instances, an additional column specifies the type of attack executed, providing valuable insights into the nature of each attack scenario.

TABLE I. DATASET DISTRIBUTION

Attack Instances	10000
Non-attack instances	90000
Total instances	100000

B. *Oversampling and Undersampling*

The ratio of attack to non-attack instances is 1:9 causing the dataset to be slightly imbalanced. To improve the performance of classification algorithms, we employed two essential techniques: Synthetic Minority Over-sampling Technique (SMOTE) and Random Under-sampling. SMOTE generates synthetic instances for the minority class by interpolating existing instances, effectively balancing class proportions [17]. This technique mitigates the risk of the model favouring the majority class due to its higher representation. Conversely, Random Under-sampling reduces the majority of class instances randomly, aligning class distributions [18]. By employing SMOTE and Random Under-sampling, we aimed to strike a harmonious balance between class representations, enabling the models to train on a more equitable dataset.

IV. CLASSIFICATION ALGORITHMS

The dataset is split into training and testing data in a ratio of 1:4. 10 different supervised learning algorithms are trained on this data comprising 17998 non-attack instances and 2002 attack instances.

A. *Supervised Classification Algorithms*

1) Logistic regression: Logistic Regression is a linear classifier that estimates instance probabilities by identifying an optimal hyperplane separating different class data points. It is particularly essential when assuming a linear relationship between input features and the outcome.

2) Support Vector Machine (SVM): Support Vector Machine aims to find the hyperplane that maximizes the margin between data points of different classes, thereby enhancing the separation between classes. SVM is particularly effective in handling complex datasets and can handle non-linear decision boundaries through kernel transformations.

3) k-Nearest Neighbours (KNN): KNN is a powerful instance-based classification algorithm that focuses on the local neighbourhood of data points. It assigns a class label to a new instance based on the majority class of its k closest neighbours. This algorithm is beneficial for capturing the local characteristics of data, making it effective for tasks with irregular data distributions and localized patterns.

4) Random forest: Random Forest is a versatile ensemble learning technique that addresses overfitting and improves predictive performance. It constructs multiple decision trees during training and combines their outputs to make predictions. By doing so, it mitigates the risk of overfitting associated with individual decision trees and provides robust results across a variety of datasets.

5) Gradient boosting: Gradient Boosting is a powerful ensemble algorithm that constructs a strong predictive model by iteratively improving upon the errors of previous iterations. It sequentially builds a series of weak learners, often decision trees, and places emphasis on instances that were misclassified earlier. This approach is particularly advantageous for capturing complex relationships and delivering high predictive accuracy.

6) Adaptive Boosting (AdaBoost): AdaBoost is a boosting algorithm that iteratively trains weak learners and combines them into a robust ensemble model. What makes AdaBoost instrumental is its ability to adapt to difficult instances by assigning higher weights to misclassified instances in the previous iteration. This adaptability enhances the model's performance and is particularly beneficial when dealing with imbalanced class distributions [5].

7) Naive bayes: Naive Bayes is a probabilistic classification algorithm that makes an instrumental simplifying assumption: feature independence given the class. Despite this assumption, it's highly efficient, making it suitable for large datasets. Naive Bayes excels in text classification and tasks where the independence assumption approximately holds, showcasing its instrumental role in such contexts.

8) Decision tree: Decision Trees are simple yet effective models that recursively partition data based on feature attributes. They're instrumental in understanding the hierarchy of decisions within a model. However, they're prone to overfitting, which can be mitigated through ensemble methods like Random Forest or Gradient Boosting, making them an essential foundational element in machine learning.

9) Multi-Layer Perceptron (MLP): Multi-Layer Perceptron is an instrumental neural network architecture composed of multiple layers of interconnected neurons. Its ability to capture complex patterns in data makes it highly versatile across various domains. However, its instrumental application requires thoughtful architecture design and hyperparameter tuning to prevent overfitting and ensure effective learning [19].

10)XGBoost: XGBoost is an advanced gradient boosting library instrumental for improved model performance. It incorporates regularization techniques, optimized tree pruning, and effective handling of missing data [20]. XGBoost's instrumental role lies in its capability to provide robust results with minimal overfitting, making it a go-to choice for boosting-based ensemble methods.

All classification algorithms have been trained on the same data with default parameters. The data has been scaled using the standard scaler. The decision function shape for the SVM algorithm is One-vs-Rest by default and max iterations for Logistic Regression have been set to 1000 to account for the size of the dataset. Now let us explore the two ensemble-based classification algorithms in consideration for this comparative analysis.

B. Ensemble-based Classification Algorithms

1) Weighted ensemble classifier: A weighted ensemble classifier is a machine learning technique that combines the predictions of multiple base classifiers, each with its own assigned weight (see Fig. 2). The weights reflect the strengths and weaknesses of each base classifier, and they determine the contribution of each classifier's prediction to the final ensemble prediction. By giving more weight to the predictions of more accurate or reliable classifiers, and less weight to those of less accurate ones, the weighted ensemble aims to improve the overall performance of the ensemble. The weights of the classifiers were decided using cross-validation f1 scores and came out to be in the range of 0.0872 for KNN to 0.1049 for AdaBoost.

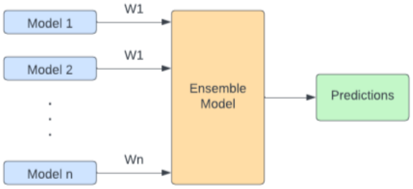


Fig. 2. Weighted ensemble classifier.

2) Majority voting ensemble classifier: The Majority Voting ensemble method is a powerful technique used to enhance the accuracy and robustness of machine learning models [9]. It involves combining the predictions of multiple individual models to make a final prediction. In Majority Voting, each model's prediction is considered a "vote" for a specific class label. The class label that receives the most votes becomes the ensemble's final prediction. This method leverages the wisdom of the crowd by aggregating the predictions of multiple models, which can lead to more accurate and reliable results. Fig. 3 diagrammatically shows its working. One of the key advantages of Majority Voting is its ability to reduce variance and errors. Even if some individual models make incorrect predictions, the ensemble can still provide accurate results if the majority of the models are correct. In cases where there is a tie in the votes, various strategies can be employed to handle it.

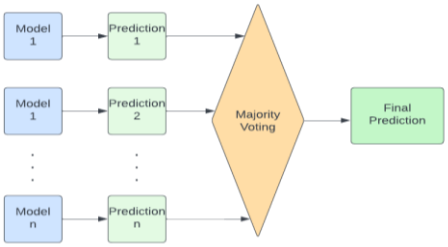


Fig. 3. Majority voting classifier.

V. Results and Discussion

In this section, we will conduct a detailed examination of the Weighted Ensemble and Majority Voting algorithms, specifically in the context of identifying potential attacks within OpenStack cloud configurations and being able to extend them to broader multi-tenant environments in the future. We will evaluate the effectiveness of these algorithms using various metrics such as accuracy, F1 score, precision, recall, and confusion matrices. By analyzing these metrics, we aim to gain insights into how well these algorithms distinguish between attack and non-attack instances in an OpenStack cloud environment. Furthermore, this section also discusses the reasoning behind choosing Majority Voting and Weighted Ensemble as the two algorithms in this comparison.

A. Performance Metrics

1) Confusion matrix: A confusion matrix provides a detailed breakdown of a model's predictions by comparing them against actual class labels. It includes four metrics: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positives represent the instances that were correctly predicted as positive by the model, while true negatives represent the instances that were correctly predicted as negative. False positives are instances incorrectly predicted as positive, and false negatives are those incorrectly predicted as negative. From these metrics, other evaluation metrics can be derived.

2) Accuracy: Accuracy is the ratio of correctly predicted instances to the total instances in a dataset, offering an overall performance assessment across all classes, particularly effective for balanced datasets. Nevertheless, in imbalanced class scenarios, accuracy might be skewed by the majority class. In such cases, metrics like precision, recall, and F1 score offer deeper insights into performance, particularly for identifying attacks in intricate multi-tenant setups.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

3) Precision: Precision evaluates the proportion of true positive predictions out of all positive predictions made by the model (see Eq. (2)). It is an important metric when the cost of false positives is high, as it indicates how trustworthy the positive predictions are. A high precision value indicates a low rate of false alarms.

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

4) Recall: Recall calculates the proportion of true positive predictions from all actual positive instances in the dataset. It is valuable when the cost of false negatives is high, as it measures how effectively the model captures all actual positives. A high recall value indicates that the model is good at identifying positives. However, a high recall value may come at the cost of more false positives.

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

5) F1 Score: The F1 score is a balanced metric that takes into account both precision and recall. It is particularly valuable when the dataset is imbalanced and there is a need to consider false positives and false negatives. The F1 score is calculated as the harmonic mean of precision and recall, helping to strike a balance between them and providing a more comprehensive assessment of a model's performance.

$$F1\ Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

B. Performance

The performance results obtained from the Weighted Ensemble and Majority Voting classifiers based on the above performance five metrics are given in Table II.

The Weighted Ensemble Classifier holds a slight performance edge over the Majority Voting Classifier in accuracy, recall, and F1 score. Although the latter has higher precision, the difference between the two precision values is not as significant as that of the two recall values. It is also noteworthy that the Majority Voting Classifier incurred an extra 1161-second runtime compared to the Weighted Ensemble. Consequently, the Weighted Ensemble Classifier emerges as the more efficient and effective choice overall.

Table III and Table IV show the confusion matrices of both classifiers for a more in-depth review of the results.

TABLE II. Results

Algorithms	**Performance Metrics**			
	Accuracy	**Precision**	**Recall**	**F1 Score**
Weighted Ensemble	0.9942	0.9876	0.9535	0.9703
Majority Voting	0.9926	0.9920	0.9336	0.9619

TABLE III. Confusion Matrix for Weighted Ensemble

Weighted Ensemble	**Actual**		
	20000 logs	**Non-attack**	**Attack**
Prediction	**Non-attack**	17974	24
	Attack	93	1909

TABLE IV. Confusion Matrix for Majority Voting

Majority Voting	**Actual**		
	20000 logs	**Non-attack**	**Attack**
Prediction	**Non-attack**	17983	15
	Attack	133	1869

TABLE V. F1 Scores after Oversampling

Algorithm with oversampling	**F1 Score**
Weighted Ensemble	0.9711
Majority Voting	**0.9702**

From these matrices, one can see that the two algorithms performed very similarly. The Weighted Ensemble could classify more attack instances correctly whereas the Majority Voting Classifier could classify more non-attack instances correctly. The Weighted Ensemble does have one disadvantage which is that it requires the calculation of weights as an extra step before the classification process. Furthermore, we have also extracted performance results after using SMOTE and RandomUndersampler on the data to get a balanced dataset [17], [18].

Table V clearly shows that oversampling reduced the performance gap between the two algorithms but the Weighted Ensemble still has slightly higher performance. Add to this, the fact that the Weighted Ensemble did not require the extra pre-processing step of Oversampling and Undersampling.

C. Discussion

Now, we delve into the evaluation of classification algorithms, and their comparison with Weighted Ensemble and Majority Voting with a particular emphasis on the stability of predictions. To simulate scenarios involving unseen data, random test-train dataset splits are utilized. A key metric used to assess the consistency of algorithmic performance is the standard deviation of F1 scores for each split.

Notably, the KNN algorithm stands out as it exhibits the lowest standard deviation (see Fig. 4), demonstrating a remarkable level of consistency across diverse dataset splits. However, its performance pales in comparison to other algorithms as shown in Fig. 5. From Fig. 4 it is evident that both the Weighted Ensemble and Majority Voting techniques prove to be the strongest contenders, displaying low standard deviations and highlighting their stability. One could argue that the Random Forest algorithm displays similar levels of stability but Fig. 5 clearly shows the lower F1 score as compared to Majority Voting and Weighted Ensemble.

This discussion provides valuable insights into the realm of intrusion detection algorithms. These algorithms carry the dual responsibility of accurately classifying threats while avoiding the pitfall of overfitting specific threat types.

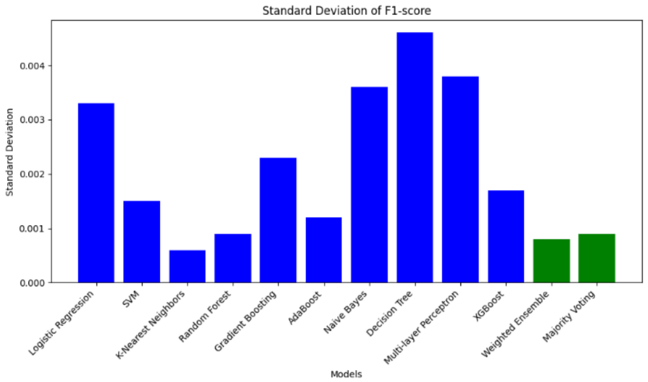


Fig. 4. Standard deviation of f1 scores of algorithms.

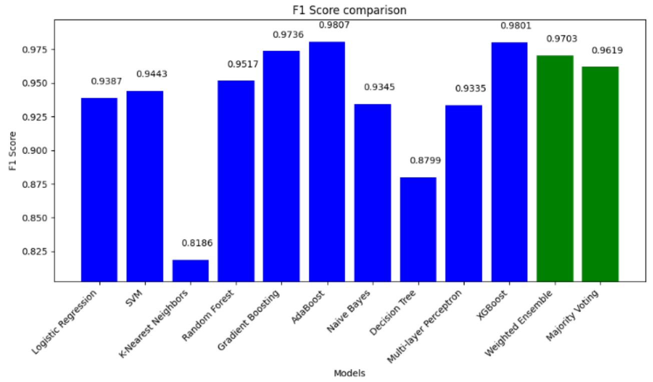


Fig. 5. Comparison of f1 scores of algorithms.

VI. Conclusion

Thus, we comprehensively compared the Weighted Ensemble and Majority Voting algorithms for intrusion detection in OpenStack cloud environments. Our analysis aimed to assess their ability to identify attacks and non-attacks and their real-world applicability. Through extensive evaluation, both algorithms demonstrated strong performance in distinguishing between attack and non-attack instances, highlighting their effectiveness as ensemble-based intrusion detection methods. While the Weighted Ensemble algorithm showcased a slight edge in terms of accuracy, recall, and F1 score, it is important to note that both algorithms demonstrated comparable performance. Additionally, the runtime analysis revealed that the Weighted Ensemble algorithm exhibited faster processing times compared to the Majority Voting algorithm, highlighting its potential efficiency advantage during real-time intrusion detection scenarios. However, it is essential to consider that real-time performance is influenced by various dynamic factors specific to the operational environment. Although one outperforms the other, both these algorithms display high performance along with stability which underscores their resilience in addressing challenges posed by unseen security threats.

Our study contributes to the ongoing discourse on enhancing cybersecurity within multi-tenant cloud environments. The findings underscore the role of ensemble techniques as valuable tools for bolstering intrusion detection capabilities. As the landscape of cyber threats evolves, future research could explore further optimizations and extensions to ensemble algorithms, aiming to refine their performance in real-world cloud environments. In conclusion, our investigation advances the understanding of ensemble-based intrusion detection, facilitating more informed decisions for ensuring the security and resilience of cloud infrastructures.

References

[1] A. Valdes, K. Skinner, Adaptive, “Model-Based Monitoring for Cyber Attack Detection”, International Workshop on Recent Advances in Intrusion Detection, Berlin, Heidelberg, 2000.

[2] J. J. Shirley and M. Priya, "A Comprehensive Survey on Ensemble Machine Learning Approaches for Detection of Intrusion in IoT Networks," 2023 International Conference on Innovations in Engineering and Technology (ICIET), Muvattupuzha, India, 2023, pp. 1-10, doi: 10.1109/ICIET57285.2023.10220795.

[3] M. Yassin, H. Ould-Slimane, C. Talhi and H. Boucheneb, "Multi-tenant intrusion detection framework as a service for SaaS," in IEEE Transactions on Services Computing, doi: 10.1109/TSC.2021.3077852.

[4] B. I. Santoso, M. R. S. Idrus and I. P. Gunawan, "Designing Network Intrusion and Detection System using signature-based method for protecting OpenStack private cloud," 2016 6th International Annual Engineering Seminar (InAES), Yogyakarta, Indonesia, 2016, pp. 61-66, doi: 10.1109/INAES.2016.7821908.

[5] Wei Hu and Weiming Hu, "Network-based intrusion detection using Adaboost algorithm," The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05), Compiegne, France, 2005, pp. 712-717, doi: 10.1109/WI.2005.107.

[6] W. Hu, W. Hu and S. Maybank, "AdaBoost-Based Algorithm for Network Intrusion Detection," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 38, no. 2, pp. 577-583, April 2008, doi: 10.1109/TSMCB.2007.914695.

[7] A. Rai, "Optimizing a New Intrusion Detection System Using Ensemble Methods and Deep Neural Network," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 527-532, doi: 10.1109/ICOEI48184.2020.9143028.

[8] P. Patil and R. Ingle, "Meta-ensemble based classifier approach for attack detection in multi-tenant distributed systems," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-6, doi: 10.1109/INCET49848.2020.9154077.

[9] Pravin Patil*, Dr. Geetanjali Kale. (2022). Stacked Anomaly Detector Guided Side Channel Attacks Detection in Multi Tenant Distributed Systems. Scandinavian Journal of Information Systems, 34(2), 17–26.

[10] E. Roponena and I. Polaka, "Classifier Selection for an Ensemble of Network Traffic Analysis Machine Learning Models," 2022 63rd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), Riga, Latvia, 2022, pp. 1-6, doi: 10.1109/ITMS56974.2022.9937116.

[11] E. Roponena and I. Polaka, "Classifier Selection for an Ensemble of Network Traffic Analysis Machine Learning Models," 2022 63rd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), Riga, Latvia, 2022, pp. 1-6, doi: 10.1109/ITMS56974.2022.9937116.

[12] V. Timčenko and S. Gajin, "Ensemble classifiers for supervised anomaly based network intrusion detection," 2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2017, pp. 13-19, doi: 10.1109/ICCP.2017.8116977.

[13] Z. Chen, W. Dong, H. Li, P. Zhang, X. Chen and J. Cao, "Collaborative network security in multi-tenant data center for cloud computing," in Tsinghua Science and Technology, vol. 19, no. 1, pp. 82-94, Feb. 2014, doi: 10.1109/TST.2014.6733211.

[14] W. H. Rankothge and S. M. N. Randeniya, "Identification and Mitigation Tool For Cross-Site Request Forgery (CSRF)," 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC), Kuching, Malaysia, 2020, pp. 1-5, doi: 10.1109/R10-HTC49770.2020.9357029.

[15] R. Shahid, S. N. K. Marwat, A. Al-Fuqaha and G. B. Brahim, "A Study of XXE Attacks Prevention Using XML Parser Configuration," 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), Al-Khobar, Saudi Arabia, 2022, pp. 830-835, doi: 10.1109/CICN56167.2022.10008276.

[16] B. W. Masduki and K. Ramli, "Improving intrusion detection system detection accuracy and reducing learning time by combining selected features selection and parameters optimization," 2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2016, pp. 397-402, doi: 10.1109/ICCSCE.2016.7893606.

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, ” SMOTE: Synthetic Minority Over-sampling Technique”, arXiv:1106.1813 [cs.AI]

[18] R. Mohammed, J. Rawashdeh and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2020, pp. 243-248, doi: 10.1109/ICICS49469.2020.239556.

[19] J. Esmaily, R. Moradinezhad and J. Ghasemi, "Intrusion detection system based on Multi-Layer Perceptron Neural Networks and Decision Tree," 2015 7th Conference on Information and Knowledge Technology (IKT), Urmia, Iran, 2015, pp. 1-5, doi: 10.1109/IKT.2015.7288736.

[20] Tianqi Chen, Carlos Guestrin, ”XGBoost: A Scalable Tree Boosting System”, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Pages 785-794.

Rural Homestay Spatial Planning and Design Based on Bert BiLSTM EIC Algorithm in the Background of Digital Ecology

Zhibin Qiu, Junghoon Mok*
Department of Space Design, Hanseo University,
Seosan 31962, Korea

***Abstract*—There is a promising development prospect in the digital ecosystem. In this context, the spatial planning of rural homestays has also received widespread attention. The research aims to better explore the advantages and determine the development direction of rural homestays, while providing two-way demand support for consumers and managers. Therefore, this study combines bidirectional long-term and short-term memory networks, pre-trained models, and emotional information attention mechanisms in deep learning. A new emotional analysis model is proposed. Then it is applied to the spatial planning of homestays near Chengdu Normal University and Chengdu Neusoft University. The experimental results show that the accuracy, recall, and F1 values of the new emotion analysis model proposed in this research reach 94%, 93%, and 94%, respectively. In terms of consumer satisfaction with the spatial location of homestays before and after the renovation, the average score of homestays near Chengdu Normal University increases by 21% compared to before the renovation. The average score of homestays near Chengdu Neusoft University increases by 40% compared to before the renovation. In summary, the new emotional analysis model proposed in this research has certain feasibility and effectiveness in the planning of rural homestay spatial location, providing new ideas for homestay spatial location planning.**

Keywords—Rural homestay; spatial planning; deep learning; emotional analysis; bidirectional long short term memory network

I. Introduction

As a key support project for the tourism industry, rural homestays have been consistently popular in recent years [1]. With the changing demand for tourism and vacation methods, the market size of the homestay industry is constantly expanding. The rise of new homestays has gradually replaced traditional farmhouses and inns [2]. Featured homestays are constantly emerging. For example, urban-rural integration and distinctive homestay planning meet personalized pursuits and experiences [3]. Abroad, homestay platforms such as Airbnb have established a huge network of accommodation resources worldwide. This specialized operational transformation further improves the quality and image of homestays [4]. In Japan, the homestay policy supports the transformation of private housing into homestays. But the operational safety and good environment of homestays need to be guaranteed [5]. However, there are currently many problems in the planning and design of homestay spaces in the Chinese market, such as unreasonable layout and inability to meet consumer needs. Inappropriate planning methods may cause a series of negative reactions.

In view of this, the research explores in the context of digital ecology and innovatively proposes the integration of Bidirectional Long ShortTerm Memory (BiLSTM), Natural Language Processing and specific elements of spatial planning. Meanwhile, considering the growth and change of the rural lodging industry, a lodging spatial location planning and design model is finally proposed. The aim is to solve the unreasonable spatial location planning in domestic homestays, provide effective advice and guidance for the spatial location planning and design of homestays, and promote the development of rural homestays. This study is divided into six sections. Section I is an introduction to the overall content. Section II outlines the related works. Section IV is about performance testing of rural homestay spatial planning model. Discussion and conclusion is given in Section V and Section VI respectively.

II. Related Works

BiLSTM is a deep learning algorithm that can effectively process sequence data. In recent years, this algorithm has achieved significant results in fields such as natural language processing and image recognition. Singla P et al. proposed an integrated model that combines wavelet transform with bidirectional long-term and short-term memory deep learning networks to predict solar irradiance at all levels within 24 hours. Wavelet transform decomposes the input time series data into different intrinsic model functions to extract its statistical features. To improve prediction accuracy, the sequences of intrinsic model functions are reduced through comprehensive experimental analysis. Wavelet decomposition components are merged. Next, each intrinsic model function subsequence is assigned a trained BiLSTM network for prediction. Finally, the predicted values of each subsequence reconstructed from the BiLSTM network are used to provide the final solar irradiance prediction at all levels. The research results indicate that the R2 value of the proposed model is 0.94. Compared to the benchmark model, prediction performance increases by 47% [6]. To predict the capacity of photovoltaic energy, Liu B et al. proposed a multi factor installed capacity prediction model based on bidirectional short-term memory grey correlation analysis. The solar photovoltaic installed capacity in China from 2020 to 2035 is predicted using this model. The research results indicate that the constructed model has high prediction accuracy. It can accurately predict China's solar photovoltaic installed capacity from 2020 to 2035, reaching 2833GW by 2035 [7]. Human activity recognition has become an important research field in human behavior analysis, human-computer interaction, and ubiquitous computing. To improve the accuracy of deep learning models in processing time series data, Challa S K et al. proposed a robust classification model for human activity recognition. This model combines convolutional neural networks and bidirectional long-term and short-term memory for design. The proposed multi branch CNN-BiLSTM network can automatically extract features from raw sensor data and minimize data preprocessing. CNN-BiLSTM can learn local features and long-term dependencies in sequence data, which helps improve the feature extraction process. The performance is evaluated using three benchmark datasets. It achieves accuracy of 96.05%, 96.37%, and 94.29% on these datasets, respectively. The experimental results show that the method combining multi-branch CNN-BiLSTM networks outperforms the other compared methods [8].

Spatial planning and design commonly referred to as "spatial design" or "spatial planning". It refers to the process of creating an indoor space that combines functionality and aesthetics in an orderly manner. Space planning and design is the allocation of appropriate functional areas for a specific environment, determining the layout and interaction of these areas. There are currently many related studies. Design space exploration provides intelligent adjustment methods for complex optimization parameters in modern advanced comprehensive space design tools. Due to the long hardware compilation time, advanced comprehensive parameter tuning is a time-consuming process. Gautier Q et al. utilized a design space exploration framework to address multiple conflicting optimization objectives and actively sought Pareto optimal solutions. The research results indicate that the adopted processing framework can achieve optimal spatial design faster. In addition, the design space exploration framework can also customize the regression model based on specific problems, thereby obtaining the model that best reflects the application design space [9]. Wu W et al. proposed a novel data-driven technology. This technology can automatically and efficiently generate floor plans for residential buildings with given boundaries. The proposed data-driven technology can mimic human spatial design processes. Firstly, the room position is determined, and then the wall position is determined, while adapting to the input building boundary. Finally, based on the spatial location of the building, a plan structure diagram is obtained. A large number of experimental results indicate that the proposed data-driven technology is effective. This technology can realistically simulate the floor plans of different floors of a building. In many cases, the floor plan generated by the research method is almost identical to the actual design plan [10].

In summary, many experts have completed a series of studies using BiLSTM, including various data prediction, action recognition, model optimization, etc [11]. For spatial planning and design problems, scholars often start with optimizing complex parameters in space. Various models are used to tune parameters to achieve the expected spatial optimization goals. For buildings like rural homestays, there is a lack of research on integrating deep learning models into the spatial planning and design of homestays. Therefore, based on the BiLSTM algorithm, an improved sentiment analysis model is proposed by combining the Bert model and emotional attention mechanism. The aim is to quickly and accurately understand customer evaluations and emotional tendencies for homestay operators, provide decision-making support, and promote the healthy development of the homestay industry.

III. Rural Homestay Spatial Planning based on Bert-BiLSTM-EIC Algorithm

There are still many problems in the current spatial planning and design of homestays, such as traffic positioning, location planning, and consumer preferences. In response to these issues, the first section constructs a new emotional analysis model through a bidirectional long-term and short-term memory network, Bert model, and emotional information attention mechanism. It effectively extracts relevant suggestions from consumers for the geographical location planning of homestays. The second section applies the model to the spatial planning of homestays near two universities, exploring the feasibility of this method.

A. Construction of a Homestay Sentiment Analysis Model Based on Bert-BiLSTM-EIC Algorithm

In recent years, the sentiment analysis method that combines Long Short Term Memory (LSTM) with deep learning has the highest popularity. LSTM updates the time record timing of cell structure through an additive calculation method, avoiding the time step feature retained when the previous state data has a significant impact on this state data, as shown in Fig. 1[12].

In Fig. 1, at time t, a value $f(t)$ between 0 and 1 is output through the internal states of $h(t-1)$ and $x(t)$, representing a trade-off between fully retained or discarded states. The activation function is the sigmoid function. The formula for the input gate is shown in Eq. (1).

$$i_t = \sigma\left(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i\right) \qquad (1)$$

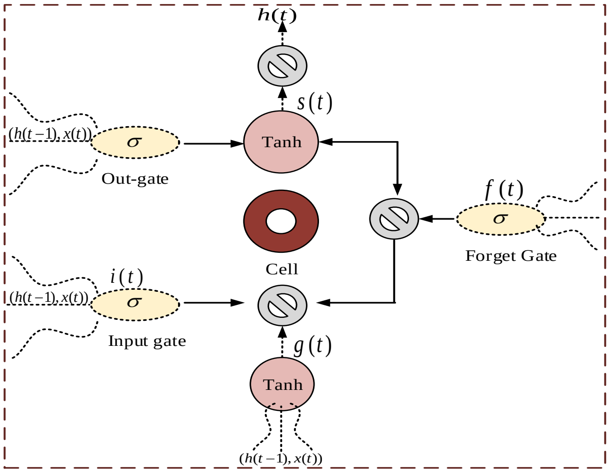


Fig. 1. LSTM network structure diagram.

In Eq. (1), i_t represents the output of the input gate. W_i and U_i represent the weight and bias of the input gate, h_{t-1} denotes the previous state. The input gate represents the information stored in $h(t-1)$ and $x(t)$ within the cell. The output gate is shown in Eq. (2).

$$o_t = \sigma\left(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o\right) \quad (2)$$

In Eq. (2), x_t denotes the output data, σ denotes the sigmoid function. W_o and U_o denote the weight and bias of the output gate, respectively, and b_o denotes the learnable parameters. The information of the output gate is usually obtained by multiplying the result of the tanh activation function with $h(t-1)$ and $x(t)$ through the sigmoid activation function [13]. The forgetting gate is shown in Eq. (3).

$$f_t = \sigma\left(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f\right) \quad (3)$$

In Eq. (3), f_t denotes the output of the forgetting gate, W_f and U_f denote the weight and bias of the forgetting gate, respectively, and b_f denotes the learnable parameter. The error function calculates the partial derivative of the weighted input of the neuron. The backpropagation error transmitted by time to the previous state is shown in Eq. (4).

$$\delta_k^T = \prod_{j-k}^{t-1} \delta_{o,t}, {}^T W_{oh} + \delta_{f,t}, {}^T W_{fh} + \delta_{i,t}, {}^T W_{ih} + \delta_{s,t}, {}^T W_{sh} \quad (4)$$

In Eq. (4), δ_k^T refers to the error of the objective function k with respect to T. W_{oh}, W_{fh}, W_{ih} and W_{sh} refer to the weights that go sequentially from unit o, f, i and s to unit h, respectively. $\delta_{o,t}$ refers to the error of the o th unit at the moment of t. $\delta_{f,t}$ refers to the error of the f th unit at the moment of t. $\delta_{i,t}$ refers to the error of the i th unit at the moment of t. $\delta_{s,t}$ denotes the error of the s th unit at the moment of t. The error formula for error propagation upwards is shown in Eq. (5).

$$\frac{\partial E}{\partial net_t^{l-1}} = (\delta_{o,t}{}^T W_{ox} + \delta_{f,t}{}^T W_{fx} + \delta_{i,t}{}^T W_{ix} + \delta_{s,t}{}^T W_{sx}) of\,'(net_t^{l-}1) \quad (5)$$

In Eq. (5), W_{ox}, W_{fx}, W_{ix} and W_{sx} refer to the weights from cell o, f, i and s to the formula cell x, respectively. The rest of the algebra is consistent with the previous explanation. However, due to the complexity and massive content of comments on homestays, relying on a single network model often leads to low efficiency and computational errors [14]. Therefore, BiLSTM with bidirectional channels is needed to process sequence data. On the basis of BiLSTM, Bert word embedding and EIC emotional information attention mechanism are introduced. The Bert-BiLSTM-EIC model is proposed. The structure diagram of the Bert-BiLSTM EIC model is shown in Fig. 2.

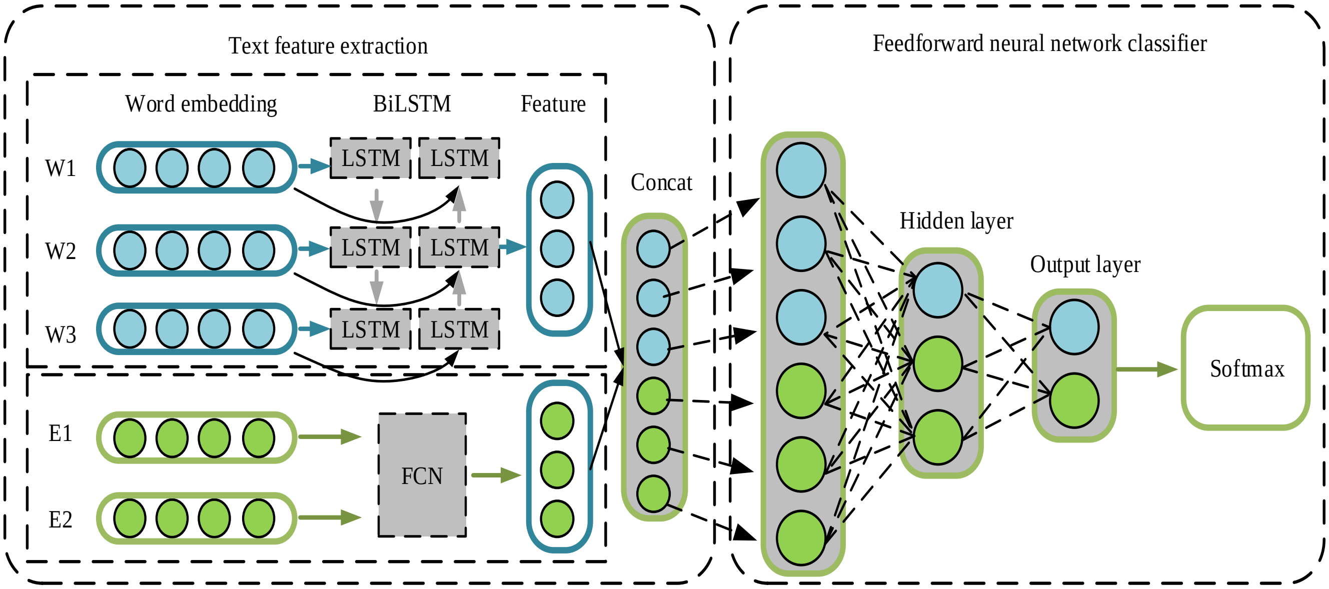


Fig. 2. Bert-BiLSTM-EIC network architecture diagram.

In Fig. 2, the Bert training model is used as the first stage. BiLSTM-EIC is used as the second stage. BiLSTM-EIC includes semantic information channels and emotional information channels. Firstly, Bert passes each input individual word through a Token embedding layer to convert each word segment into a fixed dimensional vector form. Secondly, BiLSTM-EIC extracts the emotional features of the statement, fuses the feature information and inputs it into the fully connected layer. Finally, softmax calculates the attribution probability. The Bert word embedding layer is shown in Eq. (6).

$$R_w = w_1 \oplus w_2 \oplus L \oplus w_n \tag{6}$$

In Eq. (6), R_w represents the input semantic word vector. w_1, w_2, w_3 represent different semantic participles. The word set after segmentation is used as input for the semantic channel. The Bert model obtains the semantic word input vector. The calculation of emotional channels is similar to the input of semantic channels. The calculation is shown in Eq. (7).

$$R_e = e_1 \oplus e_2 \oplus L \oplus e_m \tag{7}$$

In Eq. (7), R_e represents the input emotional information vector. e_1, e_2, e_3 represents different emotional participles. The semantic and emotional vectors are input into a network composed of forward and backward LSTM for sequence encoding to achieve information extraction and fusion, thereby improving the predictive ability of the model [15]. The calculation for model prediction is shown in Eq. (8).

$$pre = soft\max(w_o * V^{s+e} + b_o) \tag{8}$$

In Eq. (8), w_o represents the weight coefficient. pre represents the predicted emotional category. b_o represents the offset coefficient. V^{s+e} denotes the sentiment feature vector. The prediction accuracy of this model is shown in Eq. (9).

$$CV_{(k)} = \frac{1}{k}\sum_{i=1}^{k} P_i \tag{9}$$

In Eq. (9), k represents the number of datasets. i represents the combination method. P_i represents the accuracy indicator. The loss function of this model is shown in Eq. (10).

$$f_{loss} = -\frac{1}{N}\sum_{i=1}\sum_{j=1}^{M} y_i^j \log \hat{y}_i^{\,j} + \lambda\|\theta\|^2 \tag{10}$$

In Eq. (10), N is the number of samples. M is the number of emotional labels. y is the emotional label value. $\hat{y}$ is the predicted value of the model. λ is the normalization term. The preference scoring of subsequent users is shown in Eq. (11).

$$H = \begin{bmatrix} ru_1^n \, L \; ru_k^n \\ L \\ ru_1^m \, L \; ru_k^m \end{bmatrix} \tag{11}$$

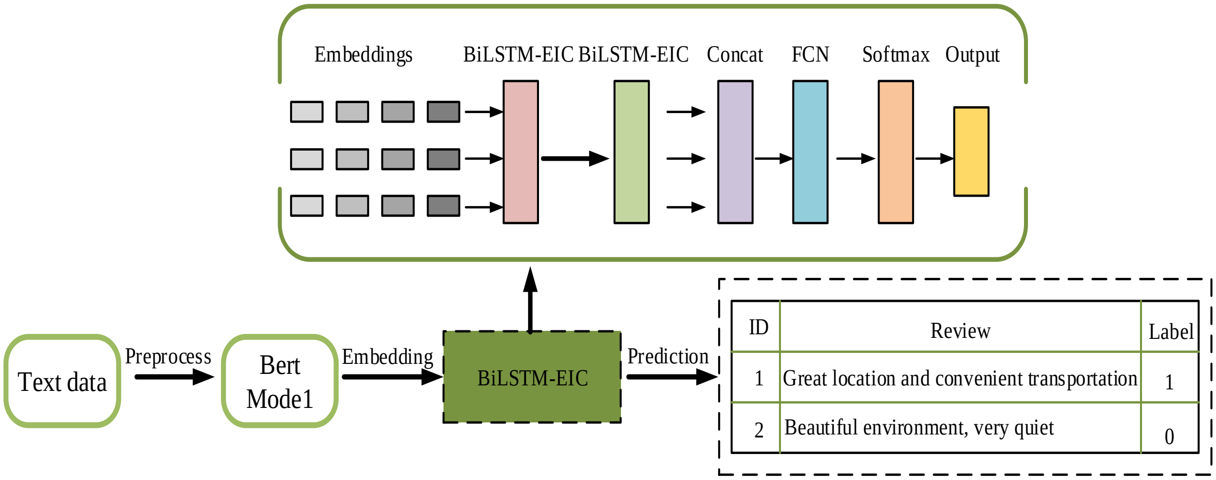
In Eq. (11), H represents the scoring set. n and m represent the regional category of homestays. k represents the number of homestays. At this point, the potential score is shown in Eq. (12).

$$R = \begin{bmatrix} r_1u_n \, L \; r_ku_n \\ K \\ r_1u_m \, L \; r_ku_m \end{bmatrix} \tag{12}$$

In Eq. (12), R represents the consumer record matrix. r_1u_n represents the number of homestay consumption in n. r_1u_m represents the number of homestay consumption in m. The location similarity of the two homestays is shown in Eq. (13).

$$similarity(h_1, h_2) = \frac{\cos(s_1, s_2 \, L \; s_3, s_4)}{e^{dis(h_1, h_2)}} \tag{13}$$

In Eq. (13), $dis(h_1, h_2)$ represents the distance between two places. s_1, s_2, s_3 and s_4 conveniently represents location, transportation, environment, and resource ratings. The sentiment analysis process combined with the BiLSTM-EIC model is shown in Fig. 3.



ID	Review	Label
1	Great location and convenient transportation	1
2	Beautiful environment, very quiet	0

Fig. 3. Emotion analysis flow of BiLSTM-EIC model.

In Fig. 3, step 1 inputs a comment dataset. Step 2 applies the Bert word embedding method to extract and embed sentences from the dataset into the tensor space. Step 3 inputs these different vectors into the BiLSTM model for semantic and feature extraction. Step 4 calculates and predicts feature information to obtain corresponding emotional results.

B. Spatial Location Planning of Rural Homestay Based on Emotional Analysis Model

After the epidemic, the national tourism industry has surged to a new height. The demand for homestays and accommodation requirements are gradually increasing [16]. According to the 2023 Homestay Industry Insight Report released by Feizhu Travel, the keywords for homestays in the hot search are shown in Table I.

In Table I, the best search terms selected by the public include geographical location, homestay environment, and price. Geographical location has become the most perplexing choice factor for consumers. As a homestay manager, the success of site selection almost determines the subsequent operation of homestays [17]. The study used GIS and market research data to identify optimal sites, with relevant location parameters including geographic location, environmental parameters and plot parameters [18]. The location parameters include 10-20 kilometers from the city center, no more than five kilometers from major tourist attractions, and close to major highways or transportation hubs. Environmental factors include noise levels not exceeding 50 dB and an Air Quality Index (AQI) of less than 100 year-round.The plot parameter refers to a slightly sloping or flat, well-drained soil area within the model's applicability range of 2,000 square meters [19]. Aiming at the above hot semantic words, this study compares the spatial locations of B&B rentals near Chengdu Teachers College and Chengdu Neusoft College by combining the proposed sentiment analysis model after the relevant parameters are determined. Chengdu Normal University is located in the university town area of Wenjiang District, Chengdu City. It has convenient subway and public transportation, and commercial districts gather, belonging to the urban area [20]. Chengdu Neusoft College is located in Qingchengshan Town, Dujiangyan Irrigation Project City. There are buses but no subways. There are fewer business districts and more mountain views. It belongs to a rural area [21]. There are nearly 20000 comments on geographical location provided by Meituan APP for homestays in two places. After randomly selecting the comments, they are input into the sentiment analysis model for feature extraction. The data analysis is shown in Fig. 4.

In Fig. 4, in the context of massive review data, consumers maintain independent opinion on the geographical location characteristics of homestays. After data processing in the emotional semantic analysis model, four main emotional analysis focuses are identified, convenient transportation, superior location, quiet environment, and rich supporting facilities. Based on the above online comment sentiment analysis data, the spatial location planning of the two homestays is designed before and after improvement, as shown in Fig. 5.

In Fig. 5(a) and Fig. 5(c) are the spatial maps of homestays before the renovation design of two universities. Fig. 5(b) and Fig. 5(d) show the spatial planning of homestays after the renovation of two universities. From Fig. 5, the original homestays of Chengdu Normal University are relatively concentrated. They are located in the downtown area. Although the transportation is convenient and shopping is convenient, the living environment and comfort are slightly poor [22]. Therefore, after the renovation, green circles represent quiet and comfortable homestays. The red circle indicates homestays with convenient transportation and superior location. Due to the unique geographical location, Chengdu Neusoft College is mostly located near mountains [23]. Quiet and comfortable homestays have become a trend. Considering the previous trend of loose distribution of homestays, and the difficulties of commercial and transportation areas, this renovation mainly focuses on bus stops and the surrounding areas of rare commercial areas. Consumers can balance both.

TABLE I. HOMESTAY HOT SEARCH KEYWORD LIST

First level keywords	Secondary keywords	Hot search words
Geographical position	Convenient transportation	Convenient
	Superior location	Shortcut
	Security environment	Secure
	Surrounding facilities	Richness
Price distribution	Below 200 yuan	Cheap
	200-500 yuan	Moderate price
	500-1000 yaun	Slightly expensive
	Above 1000 yuan	Too expensive
Residential environment	Lighting	Good indoor lighting
	Ventilate	Good indoor air
	Soundproof	Sleep undisturbed
	Water quality	Clean
Homestay quality	General homestay	Not so bad
	Boutique homestay	Cost performance
	Trendy homestay	Generally not selected
Home style	Minority	Seldom choose
	Rural	Acceptable
	Neo-chinese style	Acceptable
	Simple style	Most popular
Star standard	Five star	Don't have to
	Four-star	Maybe
	Three-star level	Consider moving in
	Two-star level	Consider moving in
	One-star level	Will not check in
Characteristics of homestay	Landscape experience	Family or couple
	Rural experience	Relax
	Artistic experience	Experience
	Sports experience	Recreation
	Cultural experience	For fun

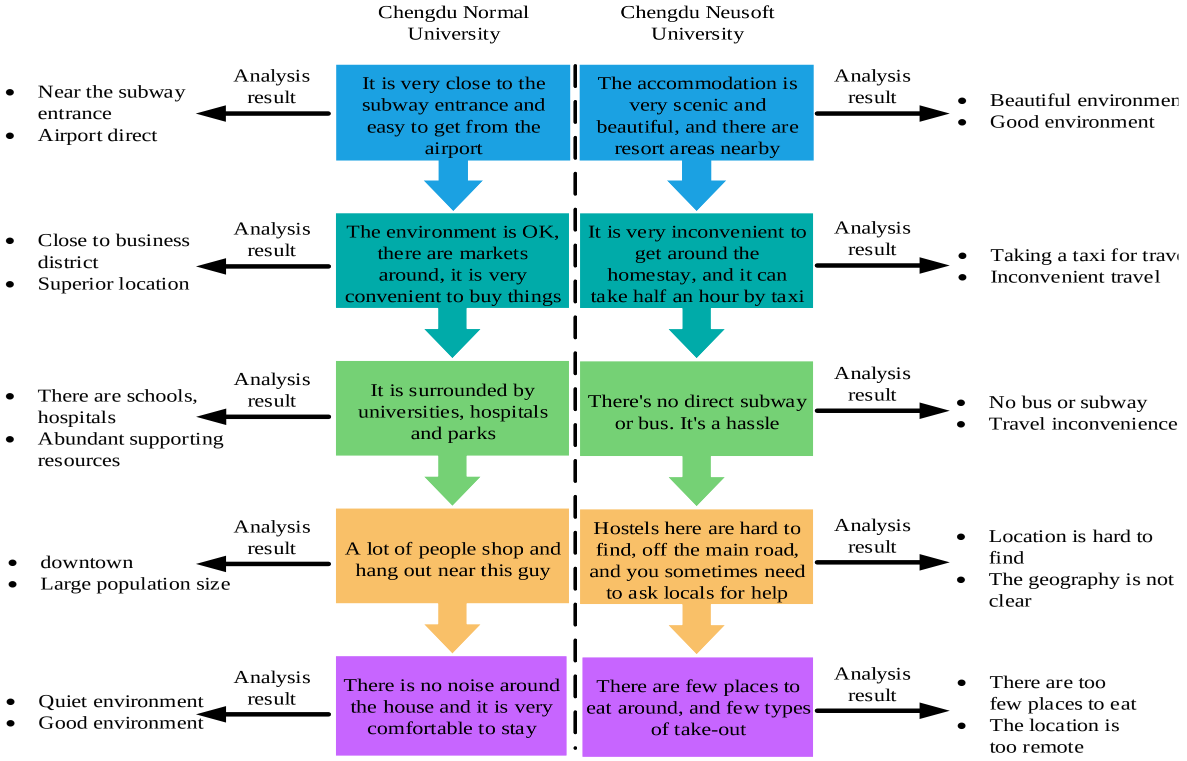
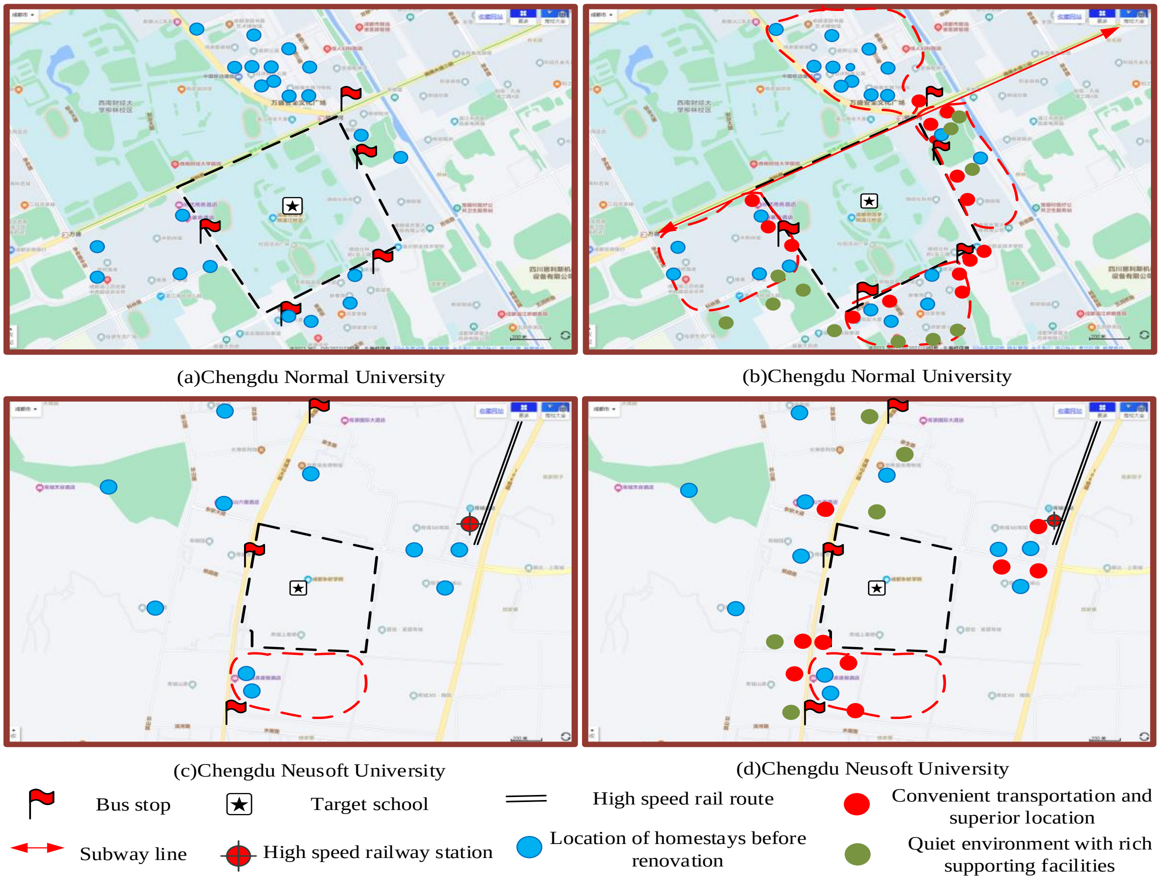

Chengdu Normal University
Chengdu Neusoft University
It is very close to the subway entrance and easy to get from the airport
Analysis result
Near the subway entrance
Airport direct
The accommodation is very scenic and beautiful, and there are resort areas nearby
Analysis result
Beautiful environmen
Good environment
The environment is OK, there are markets around, it is very convenient to buy things
Analysis result
Close to business district
Superior location
It is very inconvenient to get around the homestay, and it can take half an hour by taxi
Analysis result
Taking a taxi for trav
Inconvenient travel
It is surrounded by universities, hospitals and parks
Analysis result
There are schools, hospitals
Abundant supporting resources
There's no direct subway or bus. It's a hassle
Analysis result
No bus or subway
Travel inconvenience
A lot of people shop and hang out near this guy
Analysis result
downtown
Large population size
Hostels here are hard to find, off the main road, and you sometimes need to ask locals for help
Analysis result
Location is hard to find
The geography is not clear
There is no noise around the house and it is very comfortable to stay
Analysis result
Quiet environment
Good environment
There are few places to eat around, and few types of take-out
Analysis result
There are too few places to eat
The location is too remote

Fig. 4. Analysis of review data of homestays near two universities.


(a)Chengdu Normal University

(b)Chengdu Normal University

(c)Chengdu Neusoft University

(d)Chengdu Neusoft University

Bus stop
Target school
High speed rail route
Convenient transportation and superior location
Subway line
High speed railway station
Location of homestays before renovation
Quiet environment with rich supporting facilities

Fig. 5. Emotional analysis model of rural residential space location planning and design.

IV. Performance testing of rural homestay spatial planning model based on Bert-BiLSTM-EIC algorithm

The Bert module is used as the first part of the model training to improve its ability to be used to understand and extract contextually relevant features from the input text. This is followed by the BiLSTM layer, which is used to process the features extracted by the Bert layer to capture the long-term dependencies in the text. Finally the EIC layer is trained which is used for used for final sentiment classification based on the output of the first two layers. During the training process, the average length of the text in the training set is too long, while the amount of training data is large. Therefore, the study selectively set the Bert sequence length to 256, BiLSTM hidden layer unit to 128, and the EIC layer parameters to 64. Also, the learning rate of AdaGrade optimizer was set to 0.001, the batch size was set to 32, and the training period was set to five weeks.

Precision is the most intuitive performance metric, which indicates the ratio of instances correctly predicted by the model to the total number of instances. Recall measures the ratio of positive instances correctly identified by the model to all actual positive instances. Recall is particularly important in sentiment analysis because missing instances of positive or negative sentiment can lead to poor business decisions or misinterpreted user feedback. The F1 score is the reconciled average of accuracy and recall, which provides a balance between the two. The F1 score is particularly useful when classes are unevenly distributed or when the penalties for false positives and false negatives are the same. Compared to other metrics, these three metrics are commonly used in sentiment analysis models because they provide a comprehensive view of model performance and are easy to understand and communicate.

Firstly, a suitable experimental environment is established. Accuracy, recall, and F1 are used as evaluation indicators to test the sentiment analysis model fused with the Bert-BiLSTM-EIC algorithm. Secondly, time series, geographic similarity of homestays, and user satisfaction are used as indicators for actual performance testing.

A. *Performance Testing of Sentiment Analysis Model Based on Bert-BiLSTM-EIC Algorithm*

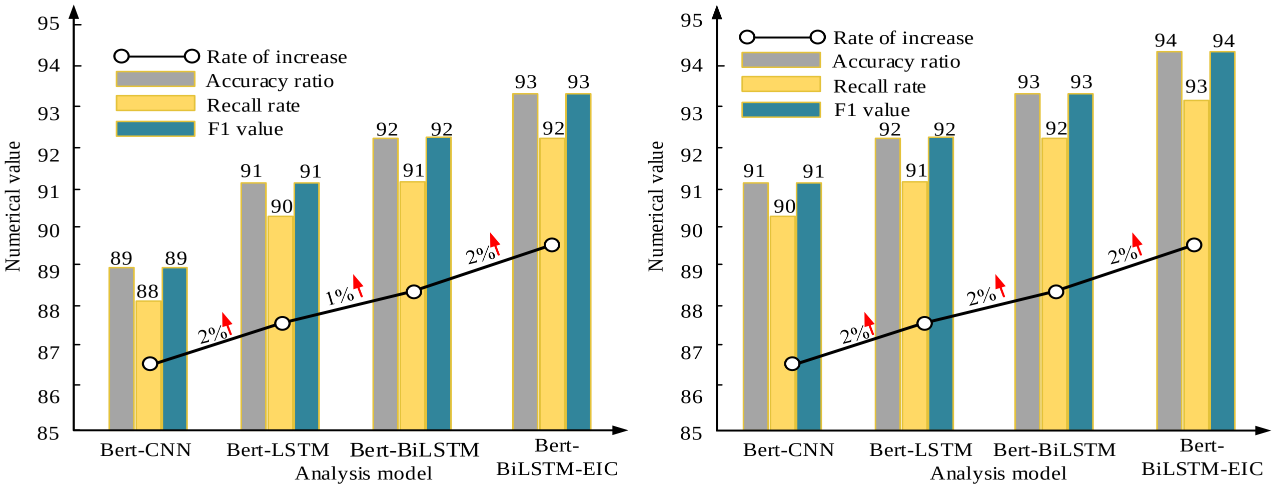
To verify the performance of the emotion analysis model proposed in this experiment, a suitable experimental environment platform is established. A guesthouse comment corpus, ChnSensiCorp, containing both Chinese and English corpora, is used as the dataset. At the same time, to ensure sufficient experimental data and the reliability of the results, a new dataset consisting of approximately 8000 positive comments and 3000 negative comments, totaling 11000, is developed by combining the review data of numerous travel softwares such as Feizhu, Ctrip, and Meituan. It is divided into a training set and a testing set according to 8:2. The specific experimental environment is shown in Table II.

Fig. 6(a) shows the accuracy, recall, and F1 values of four different analysis models in the training set. Fig. 6(b) shows the accuracy, recall, and F1 values of four different analysis models in the test set. From Fig. 6, the CNN model has the worst comprehensive experimental performance, and LSTM is good at capturing temporal features. It is slightly better than the CNN model. BiLSTM has better accuracy in capturing semantic features than LSTM. Based on the above results, the Bert-BiLSTM-EIC proposed in this study performs the best in all three aspects. Its accuracy, recall, and F1 values can reach up to 94%, 93%, and 94%, respectively. The overall score has increased by six percentage points compared to the CNN network. Therefore, the EIC emotional information attention mechanism can extract emotional word sets from text information in the set order, thereby improving the accuracy of semantic and emotional information retrieval and better capturing potential semantic features. To further verify the performance status of the Bert BiLSTM EIC analysis model proposed in this research, the loss function and accuracy are used as reference indicators during the training iteration process. The Bert-CNN model, Bert-LSTM model, Bert-BiLSTM, and Bert-BiLSTM-EIC model are used as experimental objects for testing. The specific test results are shown in Fig. 8.

Accuracy, recall, and F1 value are used as evaluation indicators. The emotion analysis model based on BiLSTM-EIC proposed in this research and three different analysis models, including Bert-CNN model, Bert-LSTM model, and Bert-BiLSTM, are compared and tested. The specific test results are shown in Fig. 6.

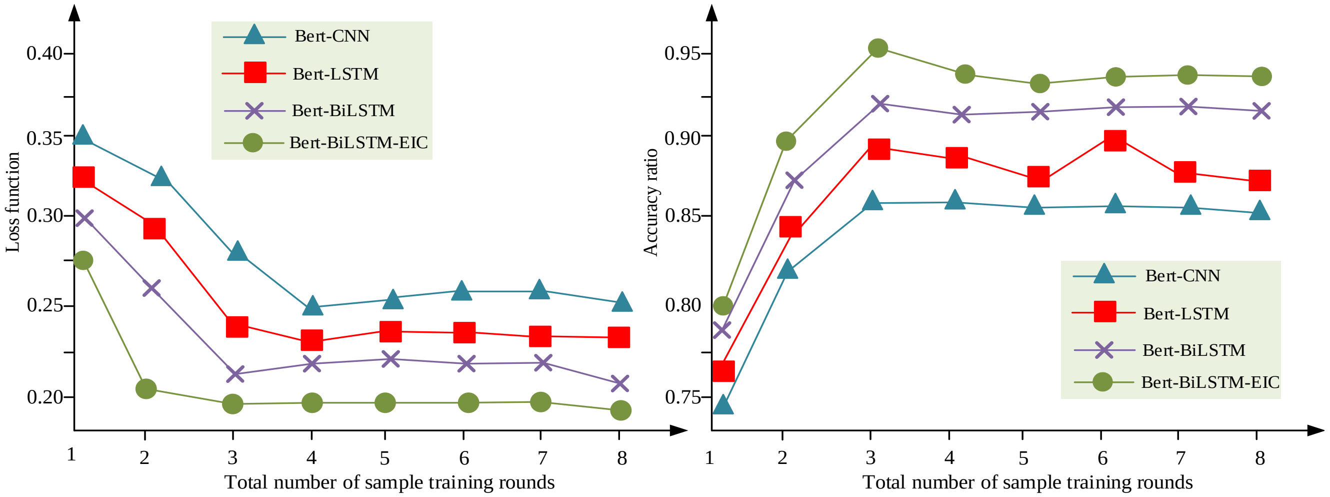
TABLE II. Experimental Environmental Parameters

Equipment environment and parameter items	Model and specific parameters
CPU	3.6GHz Intel Core i9
GPU	NVIDIA GeForce GTX 3066Ti 8G
Operating system	Windows10 64
Memory	32G
Deep learning library	TensorFlow
Amount of data for one training session	1 article
The total number of rounds required for training samples	5 rounds
Iterations	45000 times
Learning rate	0.001
Optimizer	AdaGrade



(a) Training set (b) Test set

Fig. 6. Graph of test results for four models.



(a) Loss rate comparison chart (b) Accuracy variation chart

Fig. 7. Comparison of loss function and accuracy of different models.

Fig. 7(a) shows the loss function curve for the four analytical models. Fig. 7(b) shows the accuracy curve changes of the four analysis models. In Fig. 7, the accuracy curve of the proposed Bert-BiLSTM-EIC analysis model has always been higher than the other three models. The loss rate is the lowest among the four analysis models. Throughout the entire iteration process, the performance of all four models decreases. At this point, the loss function is negatively correlated with accuracy. Although overfitting training can cause slight data fluctuations, the Bert BiLSTM-EIC analysis model tends to be more stable than the other three models. This model has a dual channel construction of semantic emotions. Therefore, the convergence speed during training is faster than the other three models.

B. *Application Test of Emotional Analysis Model in Rural Homestay Spatial Planning*

Generally speaking, the emotional consumption patterns of users are influenced by changes in time series. Therefore, based on the spatial location planning maps of Chengdu Normal University and Chengdu Neusoft University, the application effect of the sentiment analysis model in spatial planning is further demonstrated by drawing the emotional time series curves of consumers in the two places over the year. The detailed results are shown in Fig. 8.

Fig. 8(a) shows the changes in the number of reserved homestays, quiet environment homestays, and convenient transportation homestays in the vicinity of Chengdu Normal University during the year before the renovation. Fig. 8(b) shows the changes in the number of reserved homestays, quiet environment homestays, and convenient transportation homestays near Chengdu Normal University during the year after renovation. Fig. 8(c) shows the changes in the number of reserved homestays, quiet environment homestays, and convenient transportation homestays near Chengdu Neusoft College during the year before the renovation. Fig. 8(d) shows the changes in the number of reserved homestays, quiet environment homestays, and convenient transportation homestays near Chengdu Neusoft College after renovation during the year. According to Fig. 8(a) and 8(b), the demand is highest near Chengdu Normal University after the Spring Festival, from June to August, and in November. It may be due to the urgent need to rent a house after the New Year, the summer vacation from June to August for tourism, and an increase in traveling or taking the postgraduate entrance exam in November. From Fig. 8(c) and Fig. 8(d), near Chengdu Neusoft College, due to geographical location factors before the renovation, the average annual total rental volume does not fluctuate significantly. After the renovation, the total number of rented houses has increased by nearly 60. After analyzing consumer emotions, the designed homestay housing location is more favored by the public.

To better understand consumers' preferences for the geographical location of homestays, Eq. (12) is used to calculate the similarity of housing locations. The total rental volume and consumer selection evaluation of homestays after the renovation of the two places are used as the horizontal and vertical coordinates. The scatter plot of homestays in both locations is shown in Fig. 9.

Fig. 9(a) shows the location similarity of homestays near Chengdu Normal University. Fig. 9(b) shows the location similarity of homestays near Chengdu Neusoft College. From Fig. 9, public consumers near Chengdu Normal University prefer homestays with high similarity, that is, homestays with similar geographical locations. At the same time, the higher the similarity of homestay locations, the greater the rental volume is. Consumers have higher evaluation of their choices. However, near Chengdu Neusoft College, the geographical similarity of homestays is generally not high, and their geographical locations are relatively far apart.

To enable homestay operators to quickly and accurately understand customer evaluations and emotional tendencies, provide decision support, and improve service quality, the CSI (Customer Satisfaction Index) is used for evaluation. A satisfaction survey is conducted on consumers in both regions, with a score of 1-5 on the Likert scale from dissatisfied to satisfied. The data results are shown in Table III.

From Table III, the two homestays before and after the renovation have significantly improved in convenient transportation, superior location, quiet environment, and rich resource allocation. Especially in the vicinity of normal colleges, the popularity of homestays with quiet environment, convenient transportation, and superior location has generally increased. The average score increases by 21% compared to before the renovation. After planning the geographical location of homestays near the Soft College, the scores for convenient transportation and location increased significantly. The average score has increased by 40% compared to before the renovation. The proposed sentiment analysis model that integrates the Bert BiLSTM EIC algorithm has shown high feasibility in actual rural homestay spatial transformation testing. It has played a certain reference role in the subsequent spatial location planning of rural homestays.

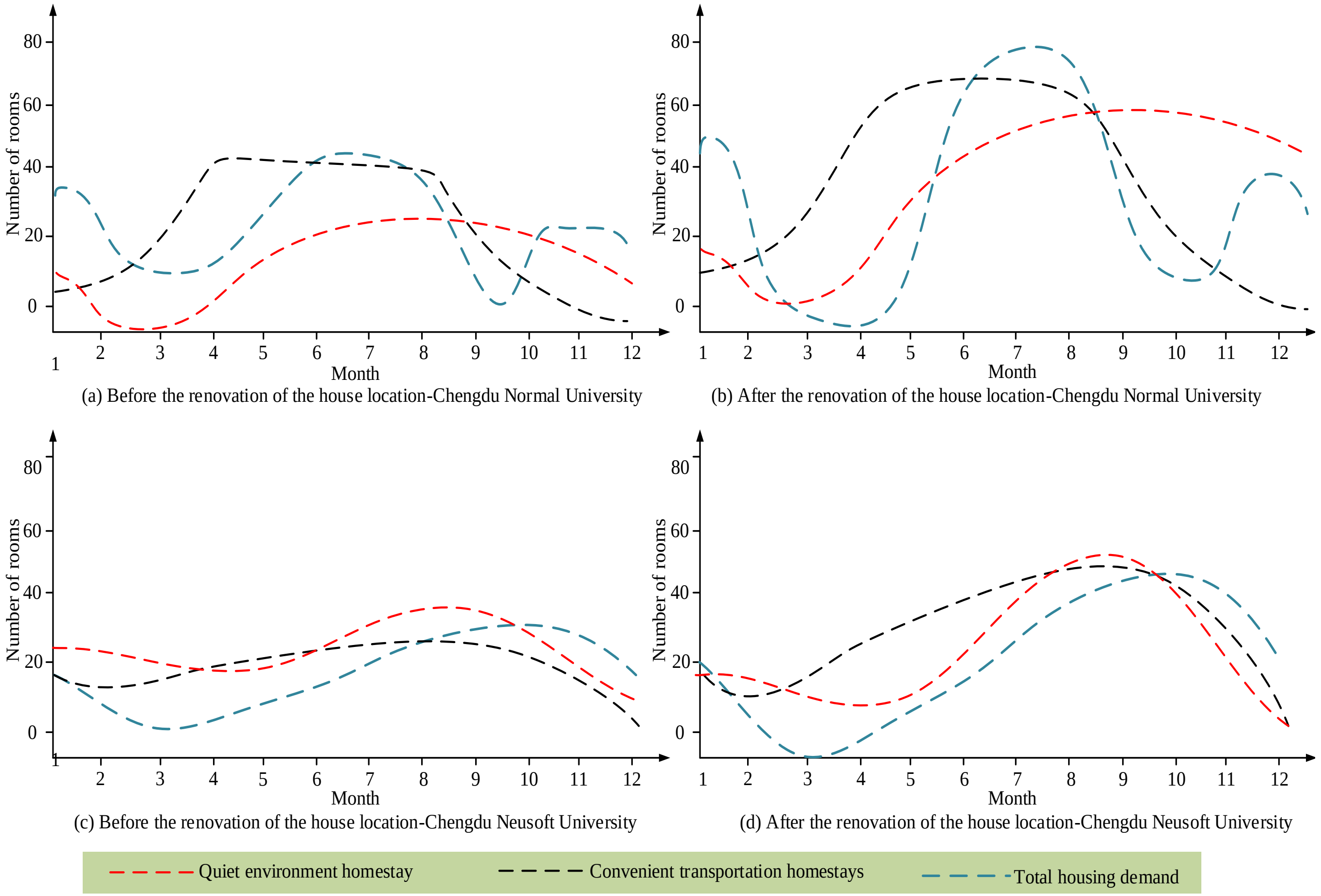
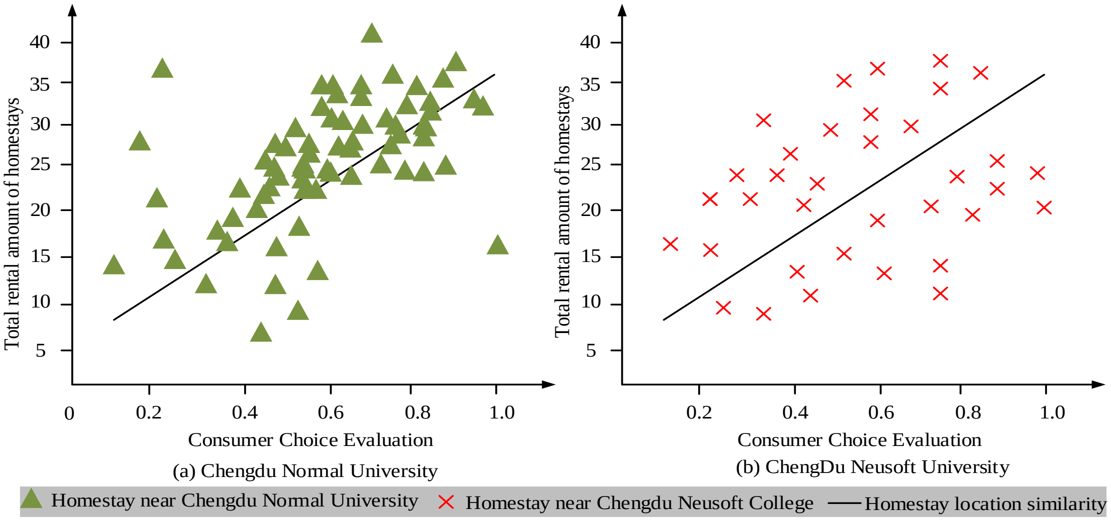


Fig. 8. The change of consumer sentiment in two places in one year.

TABLE III. CONSUMER SATISFACTION RATING

Regional variable			Convenient transportation	Superior location	Quiet environment	Abundant supporting resources	Average score
Chengdu University	Normal	Before modification	4.0	4.0	2.0	3.0	3.3
		After transformation	5.0	5.0	3.0	3.0	4.0
Chengdu University	Neusoft	Before modification	1.0	2.0	5.0	2.0	2.5
		After transformation	3.0	3.0	5.0	3.0	3.5



(a) Chengdu Normal University

(b) ChengDu Neusoft University

Fig. 9. Scatter map of homestays near two universities.

V. DISCUSSION

For this study, the data results are further expanded through the following three areas of discussion. The first is the feasibility of the algorithm effect and application. This study demonstrates the potential of digital technology in rural revitalization and tourism development by combining the Bert-BiLSTM-EIC algorithms in the spatial planning and design of rural B&Bs. The Bert module shows superiority in understanding the preferences and feedbacks of the B&B customers, while the BiLSTM performs excellently in dealing with the time-series data such as the booking trend and the flow of guests. The EIC module is introduction further optimizes the decision-making process of spatial planning, making it more accurate and efficient. Secondly, the impact of digital ecology, in the context of digital ecology, this study highlights the importance of data-driven decision making. As more and more data becomes available, it enables the public to understand the needs and challenges of the rural lodging industry more fully. The digital ecology not only provides rich data resources, but also offers new perspectives on the sustainable development of rural B&Bs through the application of algorithms. The result exists a consistent recognition with the optimal row design of neighborhood houses proposed by Teng Y et al. [24]. Finally, the innovation of spatial planning and design, the study is able to solve the complex problems in spatial planning of rural lodging more effectively by utilizing the Bert-BiLSTM-EIC algorithm. For example, the algorithm helps identify the optimal spatial layout and spatial location distribution of the B&B, while taking into account the emotional needs of different users. This approach brings innovation to traditional spatial planning and design, making it more data-driven and result-oriented. The result also has similarities with the algorithmic application in landscape classification by Zhang D et al. [25].

However, despite the positive results of this study, there are some challenges and limitations. First, access to high quality and relevant data remains a challenge. Second, the algorithmic models need to be continuously adjusted and optimized to adapt to changing market and environmental conditions. Finally, applying these techniques to rural lodging in different regions may require targeted adaptation to specific local needs and conditions. For example, sentiment analysis may be more focused on community services and quality of life feedback in rural areas, while in urban areas it may be more focused on business services and the experience of city life. Regardless of the application scenario or geographic location, the main means of studying the proposed model is to obtain key information through public and customer sentiment analysis, and then classify and identify it through algorithmic modeling, as well as give key solutions for spatial planning and positioning. In terms of expansion, the multilingual utility and scalability of the model needs to be further enhanced.

Follow-up research could explore the application of this algorithmic framework to broader areas of rural revitalization, such as agricultural production, rural education and health services. Further research on the adaptability and effectiveness of the algorithm in different cultural and geographical contexts is also necessary.

VI. Conclusion

The current spatial location planning of rural homestays lacks systematicity and scientificity. Therefore, a sentiment analysis model integrating Bert BiLSTM EIC algorithm is proposed based on deep learning networks. This model is applied to the spatial planning of homestays near Chengdu Normal University and Chengdu accuracy, recall, and F1 values of the model can reach 94%, 93%, and 94%. The comprehensive score has increased by 6% compared to the traditional CNN model. In the loss function test, the model has the lowest loss rate, the best stability, and the fastest convergence speed among the four analytical models. In the statistics of the homestay housing resources in the two regions over the year, the demand is highest after the Spring Festival, June August, and November. In the consumer satisfaction scores before and after the renovation of the two places, the average score of homestays near Chengdu Normal University increases by 21% compared to before the renovation. The average score of homestays near Chengdu Neusoft College increases by 40% compared to before the renovation. In summary, the proposed sentiment analysis model that integrates the Bert BiLSTM EIC algorithm has high accuracy and feasibility in the spatial location planning of rural homestays, which can effectively improve the spatial layout optimization and service quality of rural homestays. However, there are many limiting factors in the actual planning of homestays, such as local policies, population size, etc. Therefore, this is also an area for further improvement in subsequent experiments. For example, a detailed analysis of local policies, such as Neusoft University. For the Bert-BiLSTM-EIC algorithm, the land use regulations, building codes, and tourism promotion policies, was included in the experiment. Simulate the impact of different policy changes on B&B spatial planning in order to evaluate planning options under different policy environments. Or integrate data such as population size, population density, and population movement into the model to better understand the needs and constraints of the target area. Use the model to simulate the impact of different population dynamics on B&B planning, such as population movement during peak and off-season, and long-term population trends.

References

[1] Che L, Zhou L, Xu J. Integrating the ecosystem service in sustainable plateau spatial planning: A case study of the Yarlung Zangbo River Basin. Journal of Geographical Sciences, 2021, 31(2):281-297.

[2] Sugandi Y B W, Paturusi S A, Wiranatha A S. Community Based Homestay Management in The Village Tourism of Tete Batu, Lombok. E-Journal of Tourism, 2020, 7(2): 369-383.

[3] Guoqing Z. Study on the influence of over development and construction of traditional villages on the development of B & B and its planning strategy: A case study of Jinlin Village in Zhaoqing, Guangdong Province. Journal of Landscape Research, 2021, 13(2): 36-46.

[4] Karur K, Sharma N, Dharmatti C, Siegel J E. A survey of path planning algorithms for mobile robots[J]. Vehicles, 2021, 3(3): 448-468.

[5] Rhanoui M, Mikram M, Yousfi S, Barzali S. A CNN-BiLSTM model for document-level sentiment analysis. Machine Learning and Knowledge Extraction, 2019, 1(3): 832-847.

[6] Singla P, Duhan M, Saroha S. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. Earth Science Informatics, 2022, 15(1): 291-306.

[7] Liu B, Song C, Wang Q, Wang Y. Forecasting of China's solar PV industry installed capacity and analyzing of employment effect: based on GRA-BiLSTM model. Environmental Science and Pollution Research, 2022, 29(3): 4557-4573.

[8] Challa S K, Kumar A, Semwal V B. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. The Visual Computer, 2022, 38(12): 4095-4109.

[9] Gautier Q, Althoff A, Crutchfield C L, Kastner R. Sherlock: A multi-objective design space exploration framework. ACM Transactions on Design Automation of Electronic Systems (TODAES), 2022, 27(4): 1-20.

[10] Wu W, Fu X M, Tang R, Wang Y H, Qi Y H, Liu L G. Data-driven interior plan generation for residential buildings. ACM Transactions on Graphics (TOG), 2019, 38(6): 1-12.

[11] Roshani S, Koziel S, Yahya S I, et al. Mutual coupling reduction in antenna arrays using artificial intelligence approach and inverse neural network surrogates. Sensors, 2023, 23(16): 7089.

[12] Fang Y, Luo B, Zhao T, He D, Jiang B, Liu Q. ST-SIGMA:Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting. CAAI Transactions on Intelligence Technology, 2022, 7(4):744-757.

[13] Barlow D R, Torres L G. Planning ahead: Dynamic models forecast blue whale distribution with applications for spatial management. Journal of Applied Ecology, 2021, 58(11):2493-2504.

[14] Abdurasulovich T J. Shaxs Yuz Tasvirini Klassifikatsiyalashda Bilstm Tarmog'idan Foydalanish. Models and Methods for Increasing the Efficiency of Innovative Research, 2023, 2(24): 66-68.

[15] Dileep P, Rao K N, Bodapati P, Gokuruboyina S, Peddi R, Grover A, Sheetal A. An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm. Neural Computing and Applications, 2023, 35(10): 7253-7266.

[16] Conrow L, Mooney S, Wentz E A. The association between residential housing prices, bicycle infrastructure and ridership volumes. Urban Studies, 2021, 58(4): 787-808.

[17] Guo Y, Mustafaoglu Z, Koundal D. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. Journal of Computational and Cognitive Engineering, 2023, 2(1): 5-9.

[18] Guo P, Yuan Y, Peng Y. Analysis of Slope Stability and Disaster Law under Heavy Rainfall. Geofluids, 2021, 2021(3):1-17.DOI:10.1155/2021/5520686.

[19] Parsamehr K, Gholamalifard M, Kooch Y. Impact of Land Cover Changes on Reducing Greenhouse Emissions: Site Selection, Baseline Modeling, and Strategic Environmental Assessment of REDD plus Projects. Land Degradation and Development, 2023, 34(10):2763-2779.

[20] Ray P, Chattaraj S, Bandyopadhyay S, Jena R, Singh S, Ray S. Shifting cultivation, soil degradation and agricultural land use planning in the North-eastern hill region of India using geo-spatial techniques. Land Degradation & Development, 2021,32(14):3870-3892.

[21] Almutairi K. Determining the appropriate location for renewable hydrogen development using multi-criteria decision-making approaches. International Journal of Energy Research, 2022, 46(5):5876-5895.

[22] Joao S, Linehan D. The Colonial Hotel: spacing violence at the Grande Hotel, Beira,Mozambique. Journal of planning literature, 2023, 38(1):118-119.

[23] Che L, Zhou L, Xu J. Integrating the ecosystem service in sustainable plateau spatial planning: A case study of the Yarlung Zangbo River Basin. Journal of Geographical Sciences, 2021, 31(2):281-297.

[24] Teng Y, Yang S, Huang Y. Research on space optimization of historic blocks on Jiangnan from the perspective of place construction. Applied Mathematics and Nonlinear Sciences, 2021, 6(1): 201-210.

[25] Zhang D, Leng J, Li X. Three-Stream and Double Attention-Based DenseNet-BiLSTM for Fine Land Cover Classification of Complex Mining Landscapes. Sustainability, 2022, 14(19): 12465.

Robot Human-Machine Interaction Method Based on Natural Language Processing and Speech Recognition

Shuli Wang, Fei Long*
Foreign Language School, Harbin University of Commerce, Harbin, 150028, China

***Abstract*—With the rapid development of artificial intelligence technology, robots have gradually entered people's lives and work. The robot human-machine interaction system for image recognition has been widely used. However, there are still many problems with robot human-machine interaction methods that utilize natural language processing and speech recognition. Therefore, this study proposes a new robot human-machine interaction method that combines structured perceptron lexical analysis model and transfer dependency syntactic analysis model on the basis of existing interaction systems. The purpose is to further explore language based human-machine interaction systems and improve interaction performance. The experiment shows that the testing accuracy of the structured perceptron model reaches 95%, the recall rate reaches 81%, and the F1 value reaches 82%. The transfer dependency syntax analysis model has a data analysis speed of up to 750K/s. In simulation testing, the new robot human-machine interaction method has an accuracy of 92% compared to other existing methods, and exhibits excellent robustness and response sensitivity. In summary, research methods can provide a theoretical and practical basis for the improvement of robot interaction capabilities and the further development of human-machine collaboration.**

***Keywords*—Human-computer interaction; speech recognition; natural language processing; lexical analysis; syntactic analysis**

I. Introduction

Natural language processing is an important branch of computer science and artificial intelligence that deals with techniques that enable computers to understand, interpret and generate human language. Speech recognition, on the other hand, is the technology that converts human speech into text. With the continuous development of natural language processing and speech recognition technology, robot human-machine interaction has gradually become a research hotspot [1]. There are three common human-machine interaction methods for language robots, the earliest of which was the use of rule-based human-machine interaction methods, that is, language recognition and response through pre written rules [2]. But this method requires manual writing of a large number of rules and has poor adaptability. Then, statistical human-machine interaction methods are used to identify and understand speech inputs by establishing language and speech models [3]. But this method requires a large amount of corpus for training, which is time-consuming and not accurate. Finally, the human-machine interaction method of deep learning is utilized, which automatically extracts speech and text features through deep learning models, effectively improving the performance of language recognition [4]. With the increasing demand of people, the human-machine environment has become even more harsh, so simple deep learning models can no longer meet high requirements for completing human-machine interaction tasks. In the existing research, the design of robot human-robot interaction system that combines natural language processing technology and language recognition technology is still in the minority, and only proposes, for example, a sentiment analysis robot that can better understand the user's emotions and respond accordingly by analyzing the pitch, speed, volume of speech, and the emotional color of text. Or multimodal interaction robots provide a more natural interaction experience by analyzing the user's voice, text and body language. Despite the success of these human-robot interaction system designs in the existing market, they still face challenges such as dealing with complex and variable natural language understanding challenges and maintaining high accuracy and adaptability in diverse and dynamic environments. Therefore, the research attempts to combine natural language processing and speech recognition to propose a novel approach to human-computer interaction. The method improves the two major steps of lexical analysis and syntactic analysis, and introduces structured perceptual machine and transfer-dependent syntactic analysis for improvement respectively, to enhance the computational performance of each module, and to achieve the goal of enhancing the recognition accuracy of human-computer interaction. The study aims to explore the latest progress of natural language processing and speech recognition technologies in the field of robot human-robot interaction, analyze the limitations of existing technologies, and look forward to the future development trend. Therefore, this study attempts to combine natural language processing with speech recognition and proposes a new human-computer interaction method. This method improves the two major steps of lexical analysis and syntactic analysis to enhance the recognition accuracy of human-computer interaction. This study is divided into five sections. Section I is an introduction to the overall content of the article. Section II is an analysis and summary of research on others. Section III introduces how the improved lexical analysis model and syntactic analysis model are constructed. Section IV tests the performance of the new human-computer interaction system. Section V is a summary of the paper.

Studying the application of natural language processing and speech recognition in robot human-robot interaction is crucial for improving the level of intelligence and user experience of interaction technology. It can not only improve the efficiency of communication between people and robots, but also provide better assistive tools for specific groups (e.g., people with disabilities). In addition, this research is also

*Corresponding Author.

valuable for understanding human language and communication patterns, and can contribute to the development of knowledge in related fields. The results of the above research can reveal the efficacy and limitations of natural language processing and speech recognition technologies in different contexts and provide an empirical basis for theoretical models. For example, by analyzing the performance of robots in different linguistic contexts, the research can help us better understand the impact of language complexity on the performance of the technology. Meanwhile research findings can stimulate new research questions, such as how robots deal with dialects or non-standard languages, or how to deal with metaphors and humor in language more effectively.

II. Related Works

With the advancement of artificial intelligence technology, robots have begun to play an important role in various fields. However, one of the key challenges in making robots more intelligent and humane is how to achieve natural and smooth interaction between robots and humans. Currently, many scholars have conducted in-depth research in this field. Freire Obregon D et al., in order to further improve the recognition performance of biometric verification in human-computer interaction, used independent interest frameworks to improve the accuracy of robot audio recognition. By using high confidence image facial recognition to avoid errors caused by similarity in appearance, its accurate resolution after simulation testing is higher than traditional methods, providing a new method for robot recognition technology [5]. Ko et al. introduced a nonverbal social behavior dataset to improve the recognition and learning efficiency of robots in different scenarios. This dataset includes human body index and bone data, which robots use sensors to identify and analyze, and guide subsequent behavioral operations. The learning rate of robots under this dataset has significantly improved, and their behavior inference and response abilities have significantly improved [6]. Kim et al. conducted opinion interviews with 70 ordinary users in order to further optimize the evaluation indicators of food aid robots and improve user experience. The collection of over 500 suggestions on robot interface design and security provides more solutions for the usability, emotional value, and functional construction of food aid robots [7]. Roda Sanchez L et al. proposed an intelligent system that combines the Internet of Things and human-machine action collaboration to improve the efficiency of product manufacturing processes in the context of digital industry. This system is centered around human-computer interaction, reflecting the natural interaction between IoT inertial measurement unit equipment and robotic arms. The system meets the basic requirements of modern digital industrial manufacturing in terms of real-time performance, success rate, and acceptable level [8].

The development of speech recognition technology began in the 1950s, with initial research mainly based on spectral analysis and pattern matching of audio signals. However, due to limitations in computing power and data volume at the time, the accuracy and stability of speech recognition were not satisfactory. With the development of technology, speech recognition technology has made significant breakthroughs in many disciplinary fields. Alsayadi et al. proposed an automatic language recognition system based on convolutional neural networks to address the issue of distinguishing between Arabic language recognition techniques with and without inflections. The system is tested on a standard Arabic single speaker corpus. The results show that recognition techniques with neural networks are superior to traditional recognition techniques, reducing word error rates by 5.24% [9]. Lin's team designed a recognition technique that combines recursive neural network embedding blocks to extract advanced features in order to reduce speech loss caused by radio communication propagation. This technology integrates multi language speech recognition into a single model, thus avoiding class imbalance. The Chinese and English character error rates of this technology are 4.4% and 5.6%, respectively, which are significantly better than other methods [10]. Dong et al. proposed a significant time series method using connectionist time classification to address the issues of delayed response time and emotional noise in continuous emotional speech recognition technology. This method treats sentence labels as a chain of emotional significant events and non-emotional significant event states. This method can continuously improve the performance of emotion recognition, and when the consistency of significant emotional events is high, this improvement is more significant [11]. Yerigeri et al. proposed a mechanical and efficient speech emotion recognition technology that utilizes stress level analysis to explore the impact of stress on people's emotional changes. This technology utilizes learning algorithms to evaluate auditory and visual cues, and uses a pressure speech database for performance analysis. The overall performance of this technology is good, with an accuracy rate of 90.66% for stress related emotion recognition [12].

In summary, many academic teams have conducted extensive research in the field of robot interaction design and recognition technology, and have achieved remarkable results. Overall, the research status of robot human-machine interaction method design is constantly developing and innovating, involving the intersection and integration of multiple disciplines. This study attempts to apply natural language processing technology and speech recognition technology to the design of robot human-machine interaction, exploring how these technologies can be applied to intelligent robots to achieve more natural and convenient human-machine interaction.

III. Design of a Robot Human-Machine Interaction Method Model Combining Natural Language Processing and Speech Recognition

The natural language processing technology in robot human-machine interaction methods enables machines to understand human intentions and instructions by analyzing and understanding human language [13]. Speech recognition technology converts human speech input into text or commands, thereby achieving interaction with machines. The first section of this study will improve and innovate speech recognition technology, and the second section will improve natural language processing methods.

A. Design of Lexical Analysis Model Based on Structured Perception Machine in Natural Language Processing

The most common ways of human-computer interaction are verbal communication and behavioral communication. Speech communication involves speech recognition, while behavioral communication involves image recognition. And verbal communication is nothing more than the simplest way of interaction. The existing processing methods for natural language include lexical analysis, syntactic analysis, semantic analysis, speech recognition, and speech synthesis. Lexical analysis is the most important step in natural language processing. This step consists of three parts, namely Chinese word segmentation, part of speech tagging, and entity naming recognition. For Chinese, the results of lexical analysis will directly affect subsequent natural language processing. Perception machine is a basic binary classification algorithm proposed by American scientist Frank Rosenblatt in 1957 [14]. The goal of the perceptron is to linearly classify input data into two different categories. The perceptron model is shown in Fig. 1.

In Fig. 1, w represents the normal vector and b represents the intercept. The processing of classification problems by perceptron models is called decision boundaries. The neighborhood differentiation of this model is obvious, and the feasibility of linear implementation is high. In a space, from input to output, the perceptron model calculates the formula as shown inEq. (1).

$$f(x) = sign(w \cdot x + b) \tag{1}$$

In Eq. (1), x represents a point in the input space. b represents offset. w represents the weight value. $w \cdot x$ represents the weight of the point. $sign$ represents a symbolic function. This function is shown in Eq. (2).

$$\mathrm{si}\,gn(x) = \begin{cases} +1 \longrightarrow x \geq 0 \\ -1 \longrightarrow x < 0 \end{cases} \tag{2}$$

In Eq. (2), +1 or -1 are usually used as indicators of input and output, and the geometric interpretation of the perceptron model is shown in Eq. (3).

$$w \cdot x + b = 0 \tag{3}$$

In Eq. (3), $w \cdot x + b$ belongs to the set of all linear classification models in the feature center of the perceptual model. It corresponds to a hyperplane in the feature space. Part of speech tagging is a typical type of structured prediction. The structured prediction scoring of the perceptron model is shown in Eq. (4).

$$\hat{y} = \arg\max_{y \in Y} score_{\lambda}(x, y) \tag{4}$$

In Eq. (4), λ represents the prediction model. Y represents all selectable structures. Usually for linear models, structured perceptron is used as a training algorithm, and the classifier can assist in predicting problems such as sequence annotation. The structured prediction is shown in Eq. (5).

$$\hat{y} = \arg\max_{y \in Y} (w \cdot \phi(x, y)) \tag{5}$$

In Eq. (5), x and y represent independent variables. $\phi(x, y)$ represents a characteristic function. After the product of the new feature vector and weight points, the highest output structure is used as the decoding of the sequence annotation problem. The decoding process of this method is described in Eq. (6).

$$\delta_{t,i} = \begin{cases} w \cdot \phi(s_{0,} s_i, x_1), i = 1, \cdots, N \\ \max_{1 \leq j \leq N} (\delta_{t-1} + w \cdot \phi(s_j, s_i, x_t)), i = 1, \cdots, N \end{cases} \tag{6}$$

In Eq. (6), N represents the corresponding state. s represents the score. j belongs to any one of the state sets. t represents the time. The maximum score calculation is shown in Eq. (7).

$$S = \max_{1 \leq i \leq N} \delta_{T,i} \tag{7}$$

In Eq. (7), S represents the maximum score. T represents the corresponding time set under this score. i represents the optimal path. When there is a local optimal solution, the maximum score at this point corresponds to the label under that path, which is the i-value. In summary, this study integrates the structured perceptron model into the natural language processing of human-computer interaction, and proposes a new natural language processing framework for human-computer interaction, as shown in Fig. 2.

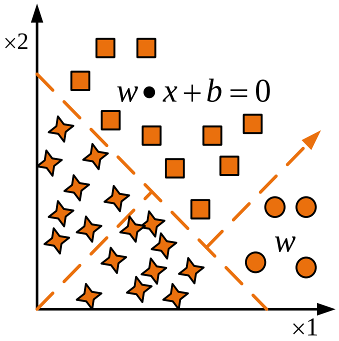


Fig. 1. Perceptron model.

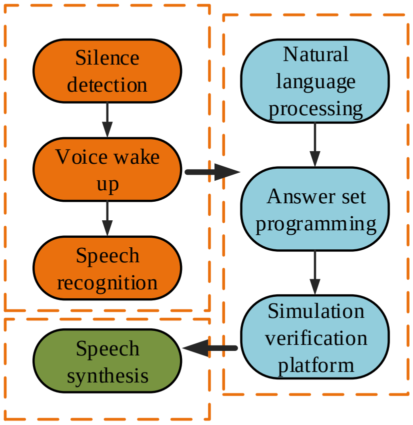


Fig. 2. Human-computer interaction natural language processing framework.

In Fig. 2, the natural language processing framework is mainly divided into three parts. Firstly, the language detection and recognition section provides information on the sounds emitted by external users. The second step is to convert the extracted sound information into linear features, which are labeled by programming data. Finally, the structure aware machine algorithm performs lexical analysis on these labeled data and converts them into communicative text.

B. Design of Key Information Speech Recognition Method Based on Dependency Syntactic Analysis

Natural language is constantly changing and cannot be represented by simple linear symbols. Meanwhile, due to the extremely complex construction of natural language, robot acquisition analysis is still too abstract. Therefore, this study continues to focus on constraint transformation of structured prediction results and remove abstraction. Natural language processing is generally manifested as the relationship between words in a sentence. Dependency syntax analysis defines this relationship as a binary nonequivalence relationship, namely the master-slave relationship [15]. To more vividly display the dependency relationships between words, the dependency tree is obtained by dividing the words in order, as shown in Fig. 3 [16].

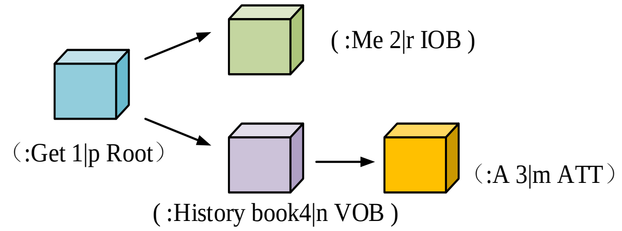


Fig. 3. "Get me a history book" depends on the syntax tree.

In Fig. 3, there are four phrase relationships in the sentence "Help me get a history book". Among them, "get" has the strongest interdependence with "me" and "history Book", while "a" has the lowest interdependence. There are two common implementation methods for dependency syntax analysis, namely the combination graph or transfer analysis method. By combining the dependency syntax analysis of graphs, using independent assumptions and establishing a model, the optimal branch solution can be found in the entire dependency tree model [17]. As shown in Eq. (8).

$$Score(x,d) = \sum_{p \subseteq d} Score_{subtree}(x, p) \tag{8}$$

In Eq. (8), w represents the weight vector. p represents a branch that conforms to the hypothesis. The analysis method of combining graphs completely depends on the maximum number of dependencies allowed in the tree. The strength of obtaining effective information in a simultaneous graph model depends on the number of features used. Under normal operation, the graph model utilizes a feature extractor to extract features for each word and transmit them to the classification scorer as scores for dependency relationships.

Based on the characteristics of dependency parsing, an improvement was made on its state machine, and conditional analysis was introduced to obtain the transfer dependency parsing algorithm [18]. The corresponding machine state under this algorithm is shown in Eq. (9).

$$s(x)_0 = ([x_0]_\sigma, [x_1 \cdots x_k]_\beta) \tag{9}$$

In Eq. (9), σ represents the stack. β represents the queue. x represents a sentence. k represents the tail element of the queue. $s(x)_0$ represents the initial state. The set of state transitions in transfer dependency syntactic analysis roughly includes move in actions, left reduction, and right reduction. The move in action is shown Eq. (10).

$$(a, x_i | \beta, A) \Rightarrow (a | x_i, \beta, A) \tag{10}$$

In Eq. (10), A represents the set of constructed dependent edges. x_i represents the state of being pushed onto the stack. Ensure that the queue is in a non-empty state and push the elements in the queue onto the stack. The calculation formula for left reduction is shown in Eq. (11).

$$([x_0 \cdots x_k, x_j, x_i], \beta, A) \Rightarrow ([x_0 \cdots x_k, x_i], \beta, A \cup (i,, j) \tag{11}$$

In Eq. (11), i and j each represent a new element. When the element in the stack is greater than 1, the two elements at the top of the stack are introduced into the set (j,i), and then the i element is re pushed onto the stack. The calculation formula for rightward reduction is shown in Eq. (12).

$$([x_0 \cdots x_k, x_j, x_i], \beta, A) \Rightarrow ([x_0 \cdots x_k, x_i], \beta, A \cup (j,i) \tag{12}$$

In Eq. (12), i and j each represent a new element. When the elements in the stack are greater than 1, the two elements at the top of the stack are introduced into the set (j,i), and then the j element is re pushed onto the stack. For example, in the short sentence "You drink water", there is a subject verb relationship between the words "you" and "drink", and a verb object relationship between "drink" and "water". Therefore, machines need two steps to establish a syntactic dependency tree during learning.

The analysis process of transfer dependency syntax analysis roughly includes transfer system, feature extraction, and action sequence transformation. The transfer system mainly covers some executable actions and their conditions. Feature extraction selects object features through manually set feature templates that combine single words, two words, and three words. The transformation of action sequences is commonly divided into static specification transfer and dynamic specification transfer [19]. To improve the persuasiveness and feasibility of machine learning, a training process for machine learning was proposed by selecting dynamic norm transfer and combining it with a structured perceptron model, as shown in Fig. 4.

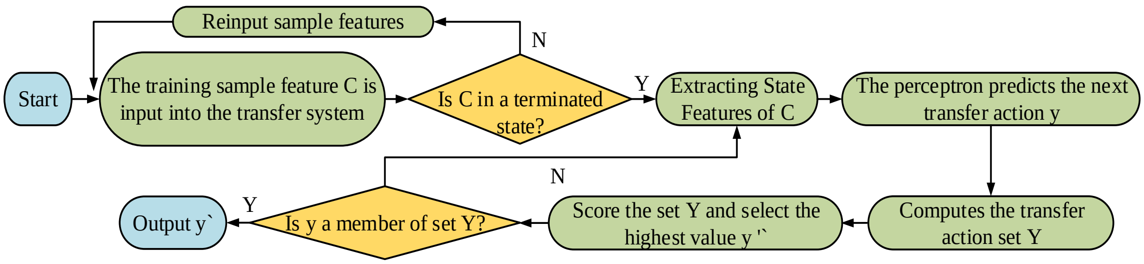

Reinput sample features
N
Start
The training sample feature C is input into the transfer system
Is C in a terminated state?
Y
Extracting State Features of C
The perceptron predicts the next transfer action y
N
Output y`
Y
Is y a member of set Y?
Score the set Y and select the highest value y '`
Computes the transfer action set Y

Fig. 4. Machine learning transfer dependency parsing training flow.

In Fig. 4, from the perspective of data structure, state judgment has been added in both the training sample input to the transfer system stage and the selection of the highest scoring feature stage, indicating that the entire process of machine learning is robust and feasible. Because after selecting features and scoring in the model, the results are usually directly output. But the process introduces attribution judgment for actions, thereby reducing the chances of self-error and increasing accuracy.

C. Construction of a Robot Human-Machine Interaction Model Combining Natural Language Processing and Speech Recognition

The first two chapters have already introduced the implementation steps of lexical analysis and syntactic analysis. This time, we will elaborate on the design of the human-machine part of the robot in the human-machine interaction model [20]. Assuming that the position of the robot is composed of an initial inertial coordinate system and a carrier reference coordinate system, the coordinate representation of the mobile robot is shown in Fig. 5.

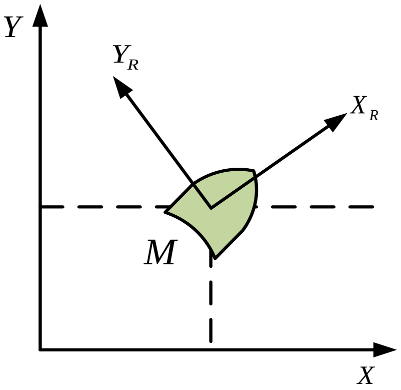


Fig. 5. Robot moving coordinate representation.

In Fig. 5, XOY represents the inertial coordinate system. $X_R MY_R$ represents the carrier coordinate system. The representation of robots in three-dimensional coordinates changes over time. If the position at time k is inferred from the position at time $k+1$, the motion model is shown in Eq. (13) [21].

$$X_r(k+1) = f(X_r(k), u(k)) \tag{13}$$

In Eq. 13, u represents the navigation weight value. f represents the state transition function. In the initial inertial coordinate system, the position vector at time $k+1$ is represented as the motion model of the robot, as shown in Eq. (14).

$$\begin{bmatrix} x(k+1) \\ y(k+1) \end{bmatrix} = \begin{bmatrix} \cos\theta - \sin\theta \\ \sin\theta, \cos\theta \end{bmatrix} \begin{bmatrix} x(k) \\ y(k) \end{bmatrix} \tag{14}$$

In Eq. (14), θ represents the heading. x and y represent the horizontal and vertical coordinates of the mobile robot, respectively. At this point, the input variable of k is $[v_r, w_r]$, where v_r represents speed. w_r represents angular velocity. After constructing the robot movement model, a new robot human-machine interaction method flow is proposed by combining the improvement scheme of natural language processing module and speech recognition module, as shown in Fig. 6 [22].

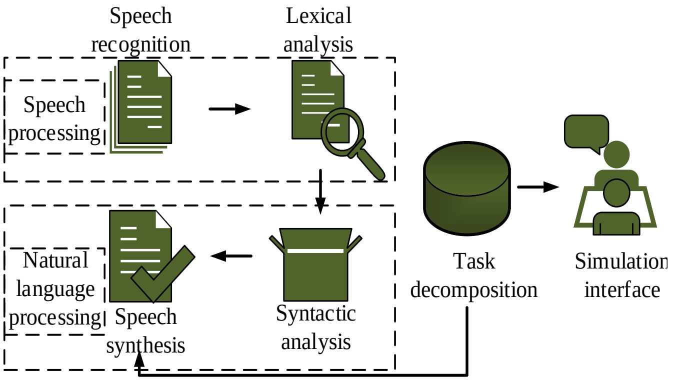


Fig. 6. New human-computer interaction process.

In Fig. 6, speech recognition and natural language processing are the two main parameter analysis modules. The speech recognition module serves as the main input part of the entire human-computer interaction process. The speech synthesis and simulation interface is the output part. In each stage, semantic recognition and synthesis are responsible for recognizing and expressing sounds, while simulation interfaces are responsible for verifying the effectiveness of interaction results [23].

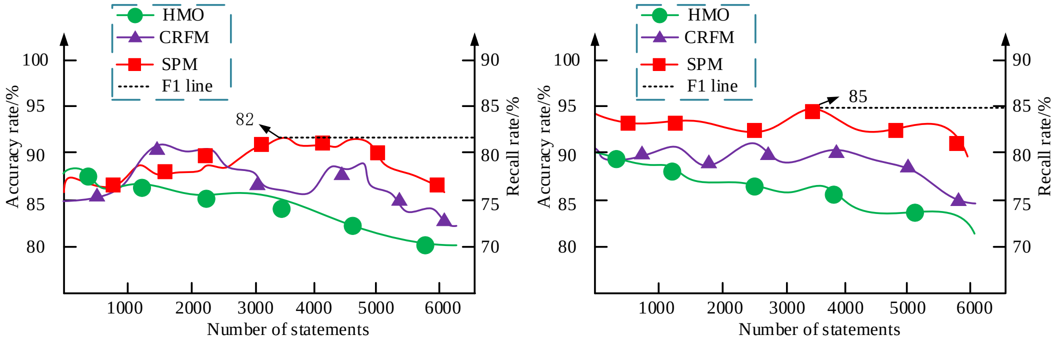
IV. Performance Testing of a Robot Human-Machine Interaction Model Combining Natural Language Processing and Speech Recognition

The first step is to establish a suitable experimental environment, set various experimental parameters and indicators, and create a reliable experimental corpus. Further extracting key feature information into robot human-machine interaction systems through natural language processing of structured perceptron models and speech recognition through transfer dependency syntax analysis. The accuracy (P), recall (R), and comprehensive evaluation index F1 value are used as evaluation indicators to conduct performance tests on the human-machine interaction model.

A. Performance Testing of Natural Language Processing Models and Speech Recognition Models

To verify the performance of the proposed new human-computer interaction model, natural language processing and speech recognition modules were tested separately. The computer system used in the experiment is Window10, the development environment is Pycharm, the language is Pyuthon3.7, and the CPU is (Intel ® Core TM i7-9700CPU@3.00GHz × 8), GPU is (NVIDIA GeForce RTX 3060 SUPER). The data source is the BCC Modern Chinese Corpus of Beijing Language and Culture University. This corpus includes over 30000 pieces of corpus information from various fields. According to an 8:2 ratio, the corpus information is divided into a training group and a testing group.

In natural language processing experiments, P value, R value, and F1 value are used as reference indicators to compare the performance of existing hidden Markov models, conditional random field models, and research models. The hidden Markov model is represented by HMO, the conditional random field model is CRFM, and the structured perceptron model is SPM. The experimental test results are shown in Fig. 7.



(a) Training set

(b) Test set

Fig. 7. Test results of different lexical analysis models.

Fig. 7 (a) and Fig. 7 (b) show the test results curves of three models in the training and testing sets. In Fig. 7, the performance of the training and testing sets of the three models shows a slow downward trend. However, the test P-values, R-values, and F1 values of the hidden Markov model perform the worst in both the training and testing sets. The combination of structured perceptron models proposed in the study performed the best, with an accuracy of up to 95%, a recall rate of up to 81%, and an F1 value of up to 82% in the test set.

Lexical analysis, as the most crucial step in natural language processing, provides the foundation for all subsequent speech processing steps. The accuracy of its analysis directly affects the efficiency of subsequent processes. Therefore, tests are conducted on the P-value, R-value, and F1 value of the three subtasks in the sequence annotation of the structured perceptron model, namely word segmentation, part of speech annotation, and named entity recognition. The results are shown in Fig. 8.

In Fig. 8, after inputting instructions from the BCC modern corpus, the average score of part of speech tagging in the structured perceptron model is the highest at 97.9%. Compared to word segmentation, its F1 value is not significantly different, but it surpasses word segmentation in terms of accuracy and recall. Due to the relatively single corpus tested, the three scores for named entity recognition are generally low.

Regarding the speech recognition process, considering that the sentences involved in human-computer interaction design in the BCC corpus are relatively fixed to be closer to life and simulate real-life communication more realistically, this study used the CTB8.0 corpus for model training and comparison. Using P value, R value, and F1 value as reference indicators, comparative tests are conducted on the probabilistic context free grammar (RCFG), semantic feature analysis (SFA), and transfer dependency syntactic analysis models. The transfer dependency parsing model is represented by TDSA, and the test results are shown in Fig. 9.

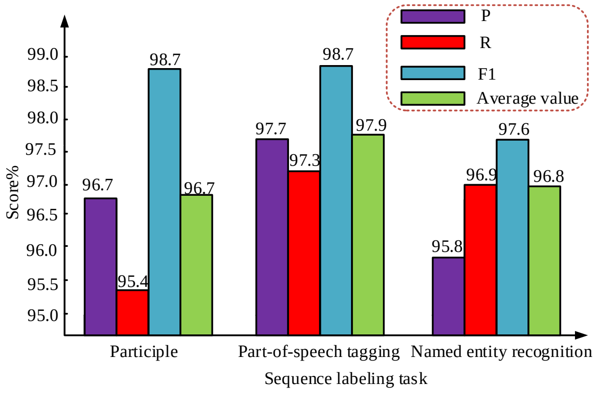
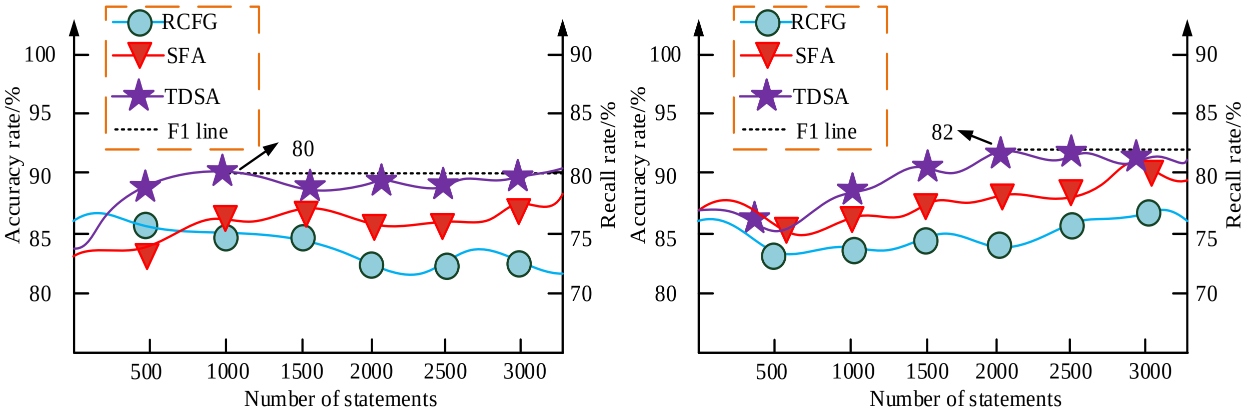


Fig. 8. Sequence annotation results of structured perceptron model.



(a) Training set (b) Test set

Fig. 9. Test results of different syntactic parsing models.

Fig. 9(a) and Fig. 9(b) show the test results of three syntactic parsing models in the training and testing sets. According to Fig. 9, in a specific corpus training environment, the accuracy of the research model is close to 90%, and compared to other models, this method has the best training results. In the test set results, the maximum P-value of the transfer dependency parsing model is close to 93%, the maximum R-value is close to 83%, and the F1 value is close to 82%.

To delve deeper into the actual performance of syntactic analysis of these three types of models, accuracy, analysis speed, and running memory are used as reference indicators, and the CTB8.0 corpus continues to be used as the test object. Table I shows the specific test results.

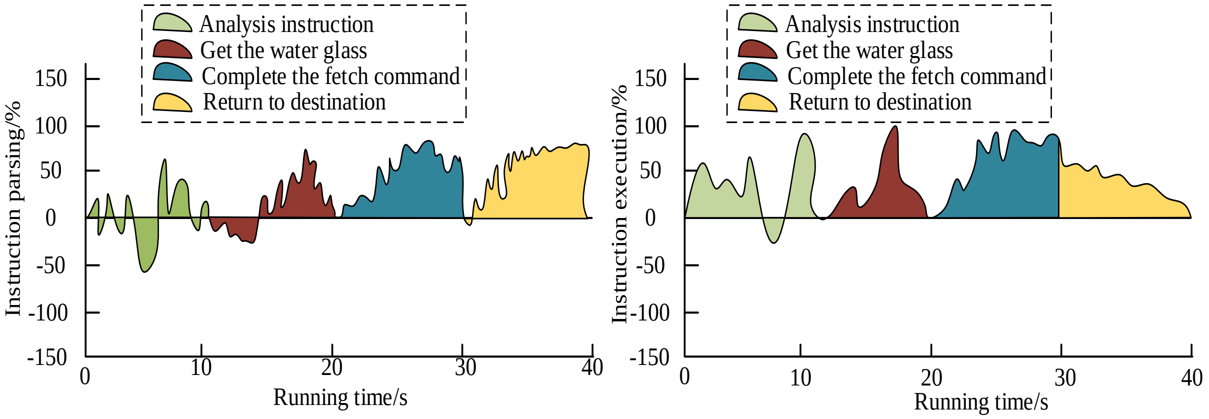
In Table I, among the three syntactic analysis models, the RCFG model has low accuracy, analysis speed, and running memory. Compared to RCFG, the accuracy and analysis speed of SFA have been improved. However, its running memory is small and not suitable for statement analysis in larger information environments. The accuracy of the research model can reach up to 96.77%, with a data analysis speed of up to 750K/s and a running memory of 126M. This can indicate that the transfer dependency syntactic analysis model proposed in this study has good practical application performance.

B. *Simulation Performance Testing of Robot Human-Machine Interaction Model*

Combining, Fig. 5, Eq. (13), and (14) for the design of the human-machine part of the robot, a human-machine interaction system using speech recognition has been built this time. It is tested using a universal Chinese textbook corpus and simulated with simple texts from daily family life. This study takes instruction parsing and execution as reference indicators, and uses the text "Give me a water cup" as the initial instruction for analysis and testing. A positive score indicates correct analysis and execution of instructions, while a negative score indicates error analysis and execution of instructions. The simulation results are shown in Fig. 10.

TABLE I. ACTUAL TEST RESULTS OF THREE SYNTACTIC ANALYSIS MODELS

Task		Accuracy/%	Speed/K/s	Memory/M
RCFG	Training set	94.31	580.00	77.00
	Test set	96.77	640.00	82.00
SFA	Training set	95.69	620.00	45.00
	Test set	96.83	650.00	68.00
TDSA	Training set	96.31	710.00	118.00
	Test set	97.28	750.00	126.00



（a）Robot command parsing situation （b）Robot command execution

Fig. 10. Robot human-computer interaction instruction execution.

Fig. 10(a) and Fig. 10(b) show the analysis accuracy and execution accuracy curves of the "Give Me a Water Cup" command executed by the human robot. In Fig. 10, the robot first performs command analysis after receiving voice commands. After disassembling the analysis results, it goes to retrieve the water cup. After completing the retrieval of the water cup, it returns to its destination. Due to the complex environment in the home environment and the presence of various interference factors in robot analysis, error analysis occurred in the instruction analysis stage for approximately two seconds. At the same time, within 10 seconds, an error action with a negative execution rate occurred. But overall, the performance during the subsequent 30 seconds of picking up the water cup and returning was relatively satisfactory.

In order to further explore the practical application performance of the robot human-machine interaction method, execution speed, execution accuracy, sensitivity, and robustness are used as reference indicators this time. This study discusses the proposed human-machine interaction method and compares it with existing robot little i human-machine interaction systems, robot Echo human-machine interaction systems, and robot Hanna human-machine interaction systems on the market. Table II shows the experimental results.

TABLE II. EXECUTION DATA OF DIFFERENT HUMAN-COMPUTER INTERACTION SYSTEMS

Human-computer interaction method	Execution time /s	Execution accuracy rate /%	Response sensitivity /%	Robustness /%
Robot little i	32	67	81	69
Robot Echo	52	88	84	54
Robot Hanna	45	74	76	83
The method proposed in this study	35	92	86	92

In Table II, the human-machine interaction execution time of little i robot is the shortest, and the execution accuracy of Echo robot is the highest. Based on the above data, it is found that the proposed human-machine interaction method combining structured perceptron and transfer dependency syntax analysis has much higher robustness and response sensitivity than other systems. Its accuracy is 92%, indicating that the combination of lexical analysis and syntactic analysis human-machine interaction method has the best execution effect in a certain speech recognition environment.

V. CONCLUSION

In order to further improve the accuracy and effectiveness of robot human-machine interaction systems, this study proposed a new method based on traditional speech recognition interaction systems. It combined a structured perceptron model and a transfer dependency syntactic analysis model for a new type of human-computer interaction. The experiment showed that the testing accuracy of the structured perceptron model in this method was as high as 95%, the recall rate was as high as 81%, and the F1 value was as high as 82%. In the testing of the transfer dependency syntactic analysis model in speech recognition, the maximum P value was close to 93%, the R value was close to 83%, and the F1 value was close to 82%. At the same time, the data analysis speed was up to 750K/s and the running memory was 126M. In the simulation testing experiment of the proposed new human-computer interaction method, although there were brief erroneous data analysis, the overall task execution rate was high. Compared to other robot human-machine interaction systems on the market, the accuracy of this method could reach 92%, and its robustness and response sensitivity were excellent. It can be seen from the above test results that, compared with the same type of language recognition interaction models, the new HCI model proposed by the study, which combines the structured perceptual machine model and the transfer dependency syntactic analysis model, is more adaptable to complex and randomly changing interaction scenarios, and the advantages of the model of this method are not only manifested in the aspects of very high accuracy, recall and F1 value, but also has an absolute leading advantage in the analysis speed and running memory. The model of this method not only shows its advantages in terms of very high accuracy, recall and F1 value, but also has absolute leading advantages in analysis speed and operation memory. Therefore, it can be said that the proposed model in the study is in the leading position in all the indexes, and can bring great impetus to the field of speech recognition human-computer interaction. In summary, the proposed new human-computer interaction method could timely and accurately respond to user instructions after correctly recognizing them, and could automatically detect and avoid erroneous voice data analysis. In addition, the high accuracy and efficiency data provided by the current study can be used as a benchmark to help future researchers optimize existing models. By analyzing and understanding the advantages of structured perceptual machines and transfer-dependent syntax, future research can build on these success factors to further improve the model. In a large number of corpus datasets, the test results of lexical analysis and syntactic analysis performed better. However, this study only focused on optimizing and improving the field of speech recognition, and had not yet introduced image analysis technology in human-computer interaction. Further research can be conducted on this basis, combined with image recognition technology, for in-depth exploration. This research provides new insights into robotic human-robot interaction systems in the field of natural language processing and speech recognition, and the results pave the way for future research. Future research can build on the current high accuracy and efficiency data for further model optimization, and delve into the causes of error data to reduce the occurrence of errors. In addition, explorations incorporating image recognition technology will open up a more comprehensive human-computer interaction experience. Cross-domain applications, such as healthcare, education or customer service, are also important directions for future research. Meanwhile, the ability to adapt to different cultural and linguistic environments, improve user experience and interaction design, and test the system's durability and application performance in long-term and real-world environments are all areas of interest. Finally, as technology evolves, research on ethical, legal, and privacy issues is indispensable to ensure the safe and responsible use of technology. Through these multiple perspectives, we are able to advance not only on the technological level, but also on the application, ethical, and legal dimensions that drive the overall development of the field.

ACKNOWLEDGMENT

The research is supported by: 2021 “Innovation” Project Support Plan for Teachers of Harbin University of Commerce. (No.20YYB020) (No.20YYC145); 2022 “Innovation” Project Support Plan for Teachers of Harbin University of Commerce. (No.22YYB252); Ministry of Education Humanities and Social Sciences Youth Foundation for Research Funds: A Study on Discourse Semantic Representation System and Resource Construction for Language Intelligence (No.23YJCZH150).

REFERENCES

[1] Qiu S, Liu Q, Zhou S, Huang W. Adversarial attack and defense technologies in natural language processing: A survey. Neurocomputing, 2022, 492(1):278-307.

[2] Peer D, Stabinger S, Engl S, Rodriguez-Sanchez A. Greedy-layer pruning: Speeding up transformer models for natural language processing. Pattern recognition letters, 2022, 157(5):76-82.

[3] Shi J, Hurdle J F, Johnson S A, Ferraro J P, Skarda D E, Finlayson S G, Samore M H, Bucher B T. Natural language processing for the surveillance of postoperative venous thromboembolism. Surgery, 2021, 170(4):1175-1182.

[4] Li Z, Ming Y, Yang L, Xue J H. Mutual-learning sequence-level knowledge distillation for automatic speech recognition. Neurocomputing, 2021, 428(7):259-267.

[5] Freire-Obregon D, Rosales-Santana K, Marin-Reyes P A Penate-Sanchez A, Lorenzo-Navarro J, Castrillon-Santana M. Improving user verification in human-robot interaction from audio or image inputs through sample quality assessment. Pattern recognition letters, 2021, 149(9):179-184.

[6] Ko W R, Jang M, Lee J, Kim J. AIR-Act2Act: Human–human interaction dataset for teaching non-verbal social behaviors to robots:. The International Journal of Robotics Research, 2021, 40(4-5):691-697.

[7] Kim H K, Jeong H, Park J, Kim W, Kim N, Park S, Park N. Development of a Comprehensive Design Guideline to Evaluate the User Experiences of Meal-Assistance Robots considering Human-Machine Social Interactions. International journal of human-computer interaction, 2022. 38(16):1687-1700.

[8] Roda-Sanchez L, Olivares T, Garrido-Hidalgo C, Luis de la Vara J, Fernandez-Caballero A. Human-robot interaction in industry 4.0 based on an internet of things real-time gesture control system. Integrated Computer-Aided Engineering, 2021, 28(2):159-175.

[9] Alsayadi H A, Abdelhamid A A, Hegazy I, Fayed Z T. Arabic speech recognition using end-toned deep learning. IET Signal Processing, 2021, 15(8):521-534.

[10] Lin Y, Yang B, Guo D, Fan P. Towards multilingual end-to-end speech recognition for air traffic control. IET intelligent transport systems, 2021, 15(9):1203-1214.

[11] Dong Y, Yang X. Affect-salient event sequence modelling for continuous speech emotion recognition. Neurocomputing, 2021, 458(11):246-258.

[12] Yerigeri V V, Ragha L K. Speech stress recognition using semi-eager learning. Cognitive Systems Research, 2021, 65(3):79-97.

[13] Bitterman D S, Miller T A, Mak R H, Savova G K. Clinical Natural Language Processing for Radiation Oncology: A Review and Practical Primer. International Journal of Radiation Oncology Biology Physics, 2021, 110(3):641-655.

[14] Ocquaye E N N, Mao Q, Xue Y, Song H. Cross lingual speech emotion recognition via triple attentive asymmetric convolutional neural network. International Journal of Intelligent Systems, 2021, 36(1):53-71.

[15] Hidayat I, Ali M Z, Arshad A. Machine Learning-Based Intrusion Detection System: An Experimental Comparison. Journal of Computational and Cognitive Engineering, 2022, 2(2):88-97.

[16] Yang B, Wang L, Wong D F. Context-Aware Self-Attention Networks for Natural Language Processing. Neurocomputing, 2021, 458(10):157-169.

[17] Mustafa H A, Al-Wesabi F N, Abdelzahir A. A Hybrid Intelligent Text Watermarking and Natural Language Processing Approach for Transferring and Receiving an Authentic English Text Via Internet. The Computer Journal, 2021,65 (2):423-435.

[18] Perboli G, Gajetti M, Fedorov S. Natural Language Processing for the identification of Human factors in aviation accidents causes: An application to the SHEL methodology. Expert Systems with Applications, 2021, 186(7):115694-115695.

[19] Jeon J H, Xu X, Zhang Y. Extraction of Construction Quality Requirements from Textual Specifications via Natural Language Processing. Transportation Research Record, 2021, 2675(9):222-237.

[20] Le T, Huang D, Apthorpe N J. SkillBot: Identifying Risky Content for Children in Alexa Skills. ACM Transactions on Internet Technology (TOIT), 2022, 22(3):79-110.

[21] Chen X, Zhang F, Zhou F, Marcello B. Multi-scale graph capsule with influence attention for information cascades prediction. International Journal of Intelligent Systems, 2022, 37(3):2584-2611.

[22] De Lope J, Grana M. An ongoing review of speech emotion recognition. Neurocomputing, 2023, 528(4):1-11.

[23] Rosenbaum T, Cohen I, Winebrand E. Differentiable Mean Opinion Score Regularization for Perceptual Speech Enhancement. Pattern recognition letters, 2023, 166(2):159-163.

Investigating of Deep Learning-based Approaches for Anomaly Detection in IoT Surveillance Systems

Jianchang HUANG*, Yakun CAI, Tingting SUN
College of Science and Technology, Hebei Agricultural University, Huanghua 061100, China

***Abstract*—Anomaly detection plays a crucial role in ensuring the security and integrity of Internet of Things (IoT) surveillance systems. Nowadays, deep learning methods have gained significant popularity in anomaly detection because of their ability to learn and extract intricate features from complex data automatically. However, despite the advancements in deep learning-based anomaly detection, several limitations and research gaps exist. These include the need for improving the interpretability of deep learning models, addressing the challenges of limited training data, handling concept drift in evolving IoT environments, and achieving real-time performance. It is crucial to conduct a comprehensive review of existing deep learning methods to address these limitations as well as identify the most accurate and effective approaches for anomaly detection in IoT surveillance systems. This review paper presents an extensive analysis of existing deep learning methods by collecting results and performance evaluations from various studies. The collected results enable the identification and comparison of the most accurate deep-learning methods for anomaly detection. Finally, the findings of this review will contribute to the development of more efficient and reliable anomaly detection techniques for enhancing the security and effectiveness of IoT surveillance systems.**

Keywords*—*Internet of Things; surveillance systems; anomaly detection; deep learning; video analysis

I. Introduction

The Internet of Things (IoT) has revolutionized various domains, including video surveillance systems, by enabling the integration of smart devices and connectivity [1, 2]. IoT video surveillance systems leverage the power of networked cameras and sensors to provide comprehensive monitoring and security solutions [3, 4]. These systems capture and process vast amounts of video data, requiring efficient techniques for analyzing and detecting anomalies in real time [5].

Video-based anomaly detection plays a vital role in IoT video surveillance systems as it enables the automatic identification of abnormal events or behaviors that deviate from expected patterns [6-8]. By leveraging computer vision algorithms, anomaly detection algorithms can detect and alert operators to potential security threats, safety violations, or irregular activities, enhancing the overall security and situational awareness of the surveillance system [9, 10].

In recent years, there have been significant advancements in video-based anomaly detection technologies. Traditional approaches relied on handcrafted features and rule-based algorithms, which often had limitations in handling complex scenarios and achieving high detection accuracy. However, with the emergence of deep learning techniques [11-13], there has been a paradigm shift in anomaly detection approaches [14]. Deep learning algorithms, such as Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs) [15], as well as Convolutional Neural Networks (CNNs) [16], have illustrated remarkable capabilities in learning discriminative representations and capturing intricate spatio-temporal patterns from video data.

Deep learning-based approaches have demonstrated superior performance in anomaly detection applications [17, 18]. They have the ability to automatically learn and extract relevant features directly from raw video data, enabling more robust and accurate anomaly detection. However, despite the promising results, there are still several research gaps and limitations that require to be addressed to exploit the potential of deep learning in this field fully.

This review paper aims to address the current limitations and research gaps in deep learning-based anomaly detection for IoT video surveillance systems. It will review and analyze the most recent methods and advancements in the field, focusing on identifying and exploring these research gaps. The paper investigates deep learning-based approaches and methodologies to tackle these challenges, aiming to enhance detection accuracy, address complex scenarios, and improve real-time performance. Additionally, extensive experimental evaluations and performance analyses will be conducted to validate the effectiveness of the suggested methods. By addressing these aspects, this review paper will contribute to the existing literature and provide valuable insights for researchers, practitioners, and system developers working on deep learning-based anomaly detection in IoT video surveillance systems.

This study delves into recent advancements in deep learning methodologies within the context of anomaly detection, examining various categories, including Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Autoencoders, Graph Convolutional Networks (GCNs), and Generative Adversarial Networks (GANs). By comprehensively exploring each deep learning category in subsequent sections, this research endeavors to not only unravel the intricacies of these methods but also to propose strategies for enhancing interpretability. Addressing the challenges posed by limited training data and concept drift, the study aims to contribute insights and methodologies that facilitate a clearer understanding of deep learning models, ensuring their effectiveness in the ever-evolving landscape of IoT environments.

The motivation behind this comprehensive review paper stems from the imperative need to address the current limitations and research gaps in deep learning-based anomaly detection for IoT video surveillance systems. By focusing on identifying anomaly in videos, the paper aims to investigate the deep learning-based approaches and methodologies that not only enhance detection accuracy but also address the challenges posed by limited training data and the dynamic nature of evolving IoT environments. The overarching goal is to improve real-time performance in complex scenarios.

The research contributions of this study are summarized as follows,

1) The review paper systematically identifies and discusses the existing research gaps and limitations in deep learning-based anomaly detection for video surveillance systems.

2) The paper introduces novel approaches that address the identified research gaps and limitations, aiming to enhance detection accuracy, address complex scenarios, and improve real-time performance in video-based anomaly detection.

3) The paper conducts extensive experimental evaluations and performance analyses to validate the effectiveness of the suggested methods, comparing them with existing state-of-the-art techniques and demonstrating their contributions regarding improved detection accuracy and real-time capabilities.

The rest of this paper is as follows, Section II review of related works. Section III discuss about research methodology. Section IV outlines the performance metrics. Section V presents results and discussion. Finally, this paper concludes in Section VI.

II. Related Work

The authors in [19] present a study on video anomaly detection with compact feature sets for online performance. The research methodology involves developing a framework that extracts compact yet discriminative features from video data to detect real-time anomalies. Key features of the study include the use of deep learning techniques for feature extraction, the incorporation of temporal information for enhanced anomaly detection, and the focus on online performance to ensure timely detection. The findings demonstrate that the suggested approach achieves efficient as well as accurate anomaly detection while reducing the computational complexity. However, one limitation highlighted in the study is the potential trade-off between the compactness of feature sets and the detection accuracy, which requires careful optimization. Overall, this research provides valuable insights into developing efficient video anomaly detection systems with compact feature sets for real-time applications.

In study [20], the application of neural networks for anomaly detection in videos is presented specifically in the context of video surveillance applications. The study presents a comprehensive overview of various neural network approaches, such as RNNs and CNNs, for analyzing video data and identifying anomalies. The findings highlight the effectiveness of neural networks in detecting anomalies in video surveillance data, showcasing their ability to capture complex spatial and temporal patterns. The paper emphasizes the potential of neural networks to enhance video surveillance systems by providing accurate and efficient anomaly detection capabilities, paving the way for improved security and monitoring in various real-world applications.

A thorough survey of deep learning-based techniques for video anomaly detection was published in study [21]. The research methodology involves an extensive examination of existing literature in the field, focusing on deep learning approaches applied to video anomaly detection. The key features of the study include categorizing and analyzing various deep learning methods, such as CNNs, RNNs, Autoencoders, GANs, and GCNs, in terms of their application, strengths, and limitations. The findings highlight the effectiveness of deep learning techniques in detecting anomalies in video data while acknowledging the challenges and limitations associated with each approach. This review serves as a valuable resource for researchers and practitioners, offering insights into the current state-of-the-art deep learning methods and their implications in video anomaly detection.

This paper in [4] focuses on anomaly detection using edge computing in video surveillance systems. The research methodology involves implementing an edge computing framework for real-time video analysis and anomaly detection. Key features of the study include utilizing edge devices to process video data locally, reducing latency and bandwidth requirements, and applying deep learning algorithms for anomaly detection. The findings demonstrate the effectiveness of edge computing in improving real-time anomaly detection performance. However, the study acknowledges limitations such as limited computational resources on edge devices and potential challenges in scaling the system. Overall, this research provides insights into leveraging edge computing for video surveillance anomaly detection while recognizing the associated limitations.

Finally, in study [22], a taxonomy of deep models for anomaly detection in surveillance videos offers a comprehensive review and performance analysis. The study systematically categorizes various deep learning models based on their thematic attributes and provides a detailed examination of each category's strengths, limitations, and performance metrics. The findings highlight the effectiveness of deep models in detecting anomalies in surveillance videos, showcasing their ability to capture intricate spatial and temporal patterns. The paper emphasizes the importance of selecting appropriate deep-learning architectures based on the specific requirements of surveillance applications. Overall, this research provides valuable insights into the state-of-the-art deep learning approaches for anomaly detection in surveillance videos, facilitating informed decision-making for implementing robust and efficient surveillance systems.

As results, the papers contribute to advancing video anomaly detection using deep learning while addressing critical challenges and needs in the field. The research in [16] emphasizes real-time performance by introducing a framework with compact feature sets, addressing the need for efficiency; however, the potential trade-off between compactness and accuracy requires careful consideration. The study in [17] contributes to improved interpretability by exploring neural networks for video surveillance, capturing complex patterns, although it does not explicitly tackle challenges related to limited training data or concept drift. The survey in [18] categorizes deep learning methods, providing a comprehensive overview but leaves room for deeper exploration of strategies for handling limited training data and concept drift. The research in [4] focuses on real-time performance through edge computing, acknowledging challenges in scalability and limited resources, indicating potential limitations. Lastly, the study in [19] offers taxonomy of deep models, aiding interpretability, but specific strategies for addressing limited training data and concept drift could be further investigated. While each paper makes notable contributions, future research should continue to bridge gaps and enhance the interpretability, handling of limited training data, addressing concept drift, and ensuring real-time performance in deep learning-based video anomaly detection systems.

III. Research Methodology

This study intends to investigate the recent deep learning methods in video-based anomaly detection methods. Various methods have been explored in different categories. These categories are RNNs, CNNs, Autoencoders, GCNs as well as GANs. The detail of each deep learning category is discussed.

The investigation delves into the inner workings of each model, scrutinizing the learned representations and features that contribute to their predictions. Techniques such as feature visualization, activation mapping, and attention mechanisms are employed to elucidate the influential aspects of input data on model outputs. Moreover, the study scrutinizes model training procedures, optimization techniques, and generalization capabilities, aiming to understand how these factors impact the interpretability of the models. By assessing robustness, handling concept drift, and employing post-hoc explanation methods like SHAP and LIME, the research aims to provide a holistic understanding of deep learning models, making them more transparent and interpretable. Through this multifaceted investigation, the study aspires to contribute valuable insights and methodologies to address the challenges posed by limited training data and the dynamic nature of evolving IoT environments, ultimately facilitating the deployment of interpretable deep learning models in practical anomaly detection scenarios.

A. Convolutional Neural Networks (CNNs)

The CNNs have emerged as a powerful deep learning technique for analyzing visual data, particularly images and videos [14, 23, 24]. They are specifically designed to capture spatial dependencies and hierarchical patterns present in visual data, making them highly effective for tasks such as image classification, object detection, and even video-based anomaly detection. In the context of anomaly detection, CNNs can learn to detect unusual patterns or events in videos, enabling the development of systems that can automatically identify anomalies or abnormal behavior in various domains, including surveillance, industrial monitoring, and healthcare.

Several existing methods leverage CNNs for video-based anomaly detection, showcasing the effectiveness of this approach. One popular approach uses spatiotemporal CNNs [25, 26], which capture temporal and spatial information by incorporating 3D convolutions [27, 28]. These models excel at detecting anomalies that involve motion or dynamic patterns. Another approach is the use of deep feature learning, where CNNs are pre-trained on large-scale image datasets and then fine-tuned for anomaly detection on video data. By leveraging pre-trained CNN models, these methods can effectively extract high-level features from videos, enabling robust anomaly detection.

B. Recurrent Neural Networks (RNNs)

The RNNs are a class of deep learning models specifically designed to process sequential data by capturing temporal dependencies [28, 29]. They have gained significant attention in various domains, consisting of video analysis, natural language processing, and speech recognition. In the context of video-based anomaly detection, RNNs have shown great promise [30]. By considering the temporal context of video sequences, RNNs can effectively capture long-term dependencies and learn complex patterns, enabling the detection of anomalies or abnormal behavior in videos.

There are several existing RNN-based methods that leverage the power of sequential modeling for video-based anomaly detection. These methods have demonstrated their effectiveness in capturing temporal patterns and detecting video anomalies. Some notable examples include Long Short-Term Memory (LSTM) [31, 32], Gated Recurrent Unit (GRU) [33, 34], and Convolutional Recurrent Neural Network (CRNN) [35].

One widely used RNN-based method for video-based anomaly detection is LSTM. The LSTM is designed to address the vanishing gradient problem that can occur in traditional RNNs [32]. By incorporating memory cells and gating mechanisms, LSTM can capture long-term dependencies and effectively learn temporal patterns. In the context of anomaly detection, LSTM models can be trained on normal video sequences and learn to forecast the next frame based on the prior frames. Anomalies can be detected by measuring the deviation between the predicted frame as well as the actual frame. LSTM has been successfully applied in diverse domains, like surveillance [36], where it has shown promising results in detecting anomalous events like abnormal behavior or unusual object movements.

C. Autoencoders

Autoencoder Networks are a class of neural networks that are designed for data compression and unsupervised learning [37, 38]. Autoencoders consist of a decoder network that reconstructs the input data from the latent representation as well as an encoder network that maps the input data into a lower-dimensional latent space [25]. This architecture enables autoencoders to learn efficient representations of the input data by capturing the most salient features. In the context of video-based anomaly detection, autoencoders can be leveraged to detect anomalies by reconstructing normal video frames accurately, as well as identifying deviations from the learned representation.

There are several existing autoencoder-based methods that have been applied to video-based anomaly detection, showcasing the effectiveness of this approach. Some notable examples include Variational Autoencoders (VAE) [39], Stacked Autoencoders (SAE) [40], and Convolutional Autoencoders (CAE) [38, 41].

Incorporating a probabilistic interpretation, variational autoencoders (VAEs) are a type of autoencoder that may provide fresh samples from the learned latent space. VAEs model the latent space as a probability distribution and learn to encode and decode data based on this distribution. In the context of anomaly detection, VAEs can be trained on normal video frames and learn to generate new frames that adhere to the learned distribution. Anomalies can be detected by measuring the reconstruction error or by evaluating the likelihood of the generated frames. VAEs have shown promising outcomes in detecting video anomalies, such as unusual activities or objects that deviate from the learned normal behavior.

Convolutional Autoencoders (CAEs) are a variant of autoencoders specifically designed for handling image and video data. CAEs utilize convolutional layers in the encoder and decoder networks to capture spatial dependencies and preserve the structure of the input data. By learning a compact representation of normal video frames, CAEs can effectively reconstruct the input frames and identify anomalies according to deviations from the learned representation. CAEs have been successfully applied in video-based anomaly detection tasks, such as detecting abnormal events or behavior in surveillance footage or industrial monitoring. The ability of CAEs to capture both local and global features from video frames makes them suitable for detecting complex anomalies that involve spatial patterns. Therefore, Autoencoder Networks offer a powerful approach to video-based anomaly detection by learning efficient representations of normal video frames and detecting deviations from the learned representation. Existing autoencoder-based methods, such as Variational Autoencoders (VAEs) and Convolutional Autoencoders (CAEs), have demonstrated their effectiveness in capturing the salient features of video data and detecting anomalies based on reconstruction errors or generated samples. These methods contribute to the advancement of video-based anomaly detection techniques as well as enable the development of intelligent systems for identifying abnormal behavior or events in various domains.

D. Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are a deep learning models class including two components: a discriminator and a generator [42, 43]. GANs are primarily known for their ability to generate realistic synthetic data that closely resembles the training data. However, GANs have also found applications in anomaly detection, including video-based anomaly detection [44]. By training GANs on normal video sequences, they can learn the underlying patterns and generate realistic frames [45]. Anomalies can be detected by measuring the deviation between the generated frames and the actual frames, thereby identifying abnormal events or behavior in videos.

Several existing GAN-based methods have been expanded for video-based anomaly detection, showcasing the effectiveness of GANs in this domain [46]. Notable examples include AnoGAN [47], Adversarial Variational Bayes (AVB) [48], and Video Anomaly GAN (VAD) [49].

AnoGAN is a GAN-based anomaly detection method that combines the power of GANs with an unsupervised learning framework. AnoGAN utilizes a generator network to generate synthetic data and a discriminator network to differentiate between generated and real data. The anomaly detection process involves finding the latent vector that generates the closest match to a given anomalous frame. By iteratively updating the latent vector, AnoGAN can generate frames that closely resemble the anomalies. AnoGAN has shown promising results in detecting anomalies in videos by effectively capturing the underlying patterns and generating synthetic anomalies for comparison.

Video Anomaly GAN (VAD) is a GAN-based method specifically designed for video-based anomaly detection. VAD employs a spatio-temporal GAN architecture to model the temporal dependencies and spatial patterns in video sequences. The generator network in VAD generates realistic video sequences, while the discriminator network distinguishes between real and generated videos. VAD utilizes the discrepancy between the generated and real videos to detect anomalies. By training on normal video sequences, VAD learns the normal patterns and can identify deviations that indicate anomalies in the video data. VAD has shown promising results in various applications, including surveillance and industrial monitoring, by effectively capturing the complex spatio-temporal dependencies in videos.

E. Graph Convolutional Networks (GCNs)

The GCNs are a neural networks class designed to process data structured as graphs[50] . GCNs extend the capabilities of traditional CNNs to handle data that exhibits complex relationships and dependencies, such as social networks, molecular structures, and video-based anomaly detection. In the field of video-based anomaly detection, GCNs can capture the spatio-temporal relationships between video frames and effectively model the interactions between different regions of interest. By leveraging the graph structure inherent in video data, GCNs enable the detection of anomalies by learning the normal behavior patterns and identifying deviations from them.

There are several existing GCN-based methods that have been utilized to video-based anomaly detection, showcasing the effectiveness of graph-based approaches in this domain. Some notable examples include Graph Convolutional Autoencoders (GCAEs), Temporal Graph Convolutional Networks (TGCNs), and Graph Convolutional Recurrent Networks (GCRNs).

Graph Convolutional Autoencoders (GCAEs) combine the power of autoencoders with graph convolutions to learn compact representations of video frames in a graph structure. GCAEs encode the video frames as nodes in a graph and leverage the connectivity information between the frames to capture their dependencies. By reconstructing the video frames from the learned latent representations, GCAEs is able to identify anomalies by measuring the deviation among the reconstructed frames as well as the actual frames. In tasks involving the detection of anomalous events or behavior in surveillance footage or traffic monitoring, GCAEs have demonstrated promising outcomes.

Temporal Graph Convolutional Networks (TGCNs) extend the capabilities of GCNs by incorporating the temporal dynamics of video data. TGCNs model the video frames as nodes in a temporal graph, where the edges capture the temporal dependencies between frames. By performing graph convolutions across the temporal dimension, TGCNs can effectively capture the spatio-temporal patterns and dependencies in videos. TGCNs have shown great potential in detecting anomalies in video sequences, such as identifying abnormal motion patterns or unusual temporal behaviors.

The GCRNs combine the strengths of both recurrent neural networks (RNNs) and GCNs to capture both spatial and temporal dependencies in video data. GCRNs model the video frames as nodes in a graph and utilize recurrent connections to capture the temporal dynamics. By performing graph convolutions and recurrent computations, GCRNs can effectively capture the complex spatio-temporal patterns and dependencies in videos. GCRNs have demonstrated promising results in video-based anomaly detection tasks, such as detecting anomalous events or behaviors in surveillance videos or monitoring industrial processes.

F. Algorithms Hyperparameter Setting

In this study, a CUHK Avenue dataset[1] is used the video-anomaly detection experiments. Moreover, the hyperparameter setting for the algorithms are as, for RNNs, the hyperparameter setting for RNNs is: hidden size = 256, learning rate = 0.001, batch size = 16, dropout rate = 0.2, number of epochs = 501. The hyperparameter setting for CNNs is: filter size = 3x3, number of filters = 64, learning rate = 0.0001, batch size = 32, dropout rate = 0.5, number of epochs = 100. For Autoencoders, the hyperparameter setting for Autoencoders is: latent dimension = 128, learning rate = 0.0005, batch size = 64, dropout rate = 0.1, number of epochs = 2003. For GCNs, the hyperparameter setting for GCNs is: number of layers = 3, hidden size = 64, learning rate = 0.01, batch size = 128, dropout rate = 0.2, number of epochs = 300. Finally, the hyperparameter setting for GANs is: latent dimension = 256, learning rate = 0.0001, batch size = 16, dropout rate = 0.3, number of epochs = 200.

IV. Performance Metrics

Performance measurements play an essential role in evaluating the effectiveness of deep learning-based anomaly detection models. When it comes to assessing the performance of such models, three commonly used metrics are F-score, recall, and precision. These metrics aid in quantifying the model's accuracy in detecting abnormalities and offer insights into many facets of model performance.

Precision is a measure of how many of the instances labeled as anomalies by the model are actually true anomalies. It represents the true positive predictions ratio (correctly detected anomalies) to the total number of predicted anomalies (both false positives and true positives). A high precision score means that the model is more accurate at correctly identifying abnormalities and has a lower rate of false alarms. Precision is calculated using the formula Recall, as well known as sensitivity or true positive rate, is a measure of how many true anomalies the model can successfully detect. It denotes the true positive predictions ratio to the total number of actual anomalies in the dataset. F-score, also called the F1 score, is a harmonic mean of precision and recall. It supplies a single metric that balances both recall and precision, taking into account false negatives and false positives. The F-score combines recall and precision into a single value and is useful when there is a trade-off among recall and precision. The F-score is calculated utilizing the formula:

$$F - score = 2 * ((Precision * Recall) / (Precision + Recall))$$

The F-score ranges from 0 to 1, with 1 being the ideal score that indicates perfect precision and recall.

V. Results and Discussion

A. Analysis of CNN-based Methods

The table shows the recall, F-score, and precision for various CNN-based methods used in anomaly detection. We selected most used CNN methods in literature. These methods include ConvLSTM, Temporal Convolutional Network (TCN), 3D Convolutional Networks, I3D (Inflated 3D Convolutional Networks), TSN (Temporal Segment Networks), and C3D (Convolutional 3D). Fig. 1 shows the result of CNN-based methods.

Examining the precision values, we observe a range from 0.86 to 0.92. Higher precision values indicate a lower rate of false positives, reflecting the ability of the models to accurately identify anomalies while minimizing incorrect detections. The method with the highest precision in the table is 3D Convolutional Networks, suggesting a stronger precision performance in anomaly detection.

In terms of recall, the values range from 0.82 to 0.92. Recall measures the ability of the models to capture the actual anomalies present in the data. A higher recall value indicates a higher proportion of correctly identified anomalies, reducing the risk of false negatives. The I3D stands out with the highest recall score, implying its effectiveness in capturing a larger number of true anomalies.

The F-scores in the table range from 0.84 to 0.92. The F-score combines recall and precision, supplying an overall assessment of model performance. A higher F-score shows a better balance among recall and precision. In this case, I3D (Inflated 3D Convolutional Networks) demonstrates the highest F-score, indicating its effectiveness in achieving a trade-off between accurately identifying anomalies and minimizing false alarms.

[1] https://paperswithcode.com/dataset/chuk-avenue

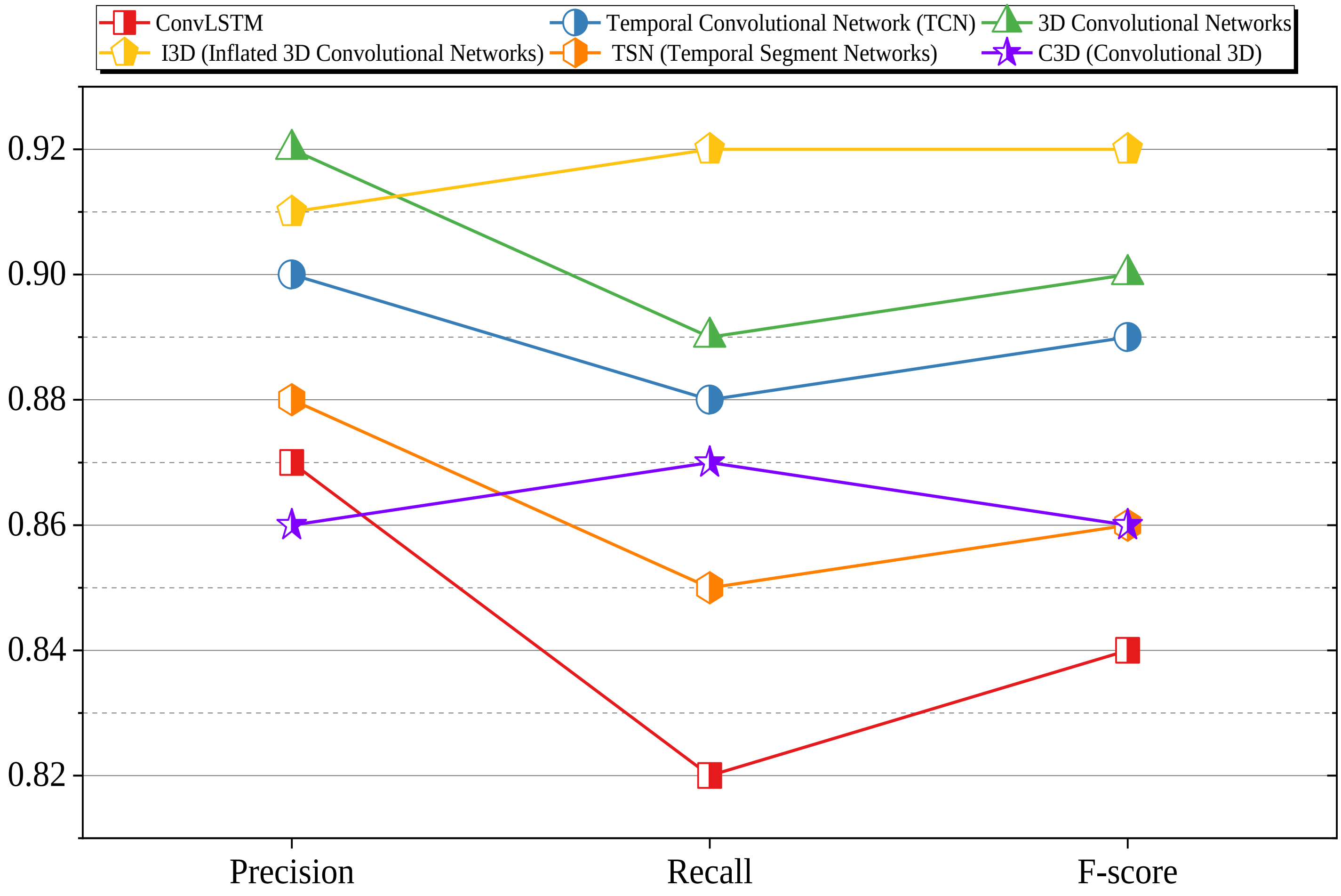


Fig. 1. Result of CNN-based methods.

B. *Analysis of RNN-based Methods*

This section presents an overview of performance metrics for different RNN-based anomaly detection methods. These methods have been evaluated using precision, recall, and F-score, providing insights into their effectiveness in detecting anomalies. Among the evaluated RNN-based methods, notable approaches include the LSTM-Autoencoder, GRU-Autoencoder, Variational LSTM (VLSTM), Temporal Convolutional LSTM (TCLSTM), Stacked LSTM, and Gated Recurrent Unit (GRU).

The performance metrics in Fig. 2 provides valuable insight into the strengths and capabilities of these RNN-based methods. Precision values in the 0.80 to 0.95 range show the ability of the models to accurately identify anomalies while minimizing false positives. This is crucial for ensuring that detected anomalies are truly meaningful and actionable. Recall values, ranging from 0.86 to 0.92, reflect the models' ability to capture a significant proportion of actual anomalies present in the data. A higher recall value indicates a reduced risk of false negatives, ensuring that fewer anomalies go undetected.

C. *Analysis of Autoencoders Methods*

This section presents result of analysis for a collection of recent Autoencoders-based anomaly detection methods, along with their corresponding precision, recall, and F-score performance metrics. We selected most cited Autoencoders methods as methods include Variational Autoencoder (VAE), Adversarial Autoencoder (AAE), Deep Autoencoder, Denoising Autoencoder, Sparse Autoencoder, and Variational Graph Autoencoder (VGAE).

As shown in Fig. 3, in terms of recall, the Sparse Autoencoder demonstrates the highest value at 0.92. This implies that the Sparse Autoencoder has a superior capability to capture a larger proportion of actual anomalies present in the data. Considering the F-score, which combines both precision and recall, the Deep Autoencoder still emerges as the method with the highest score at 0.90. This indicates that the Deep Autoencoder achieves a better balance between accurately identifying anomalies and minimizing false alarms compared to the other methods.

The better performance of the Deep Autoencoder can be attributed to its ability to learn deep, hierarchical representations of the input data. The deeper architecture permits the model to capture more complex patterns and anomalies in the data, leading to ameliorated precision, recall, and overall F-score. The Dense Autoencoder's superior performance showcases the importance of utilizing deep architectures in Autoencoders-based anomaly detection methods.

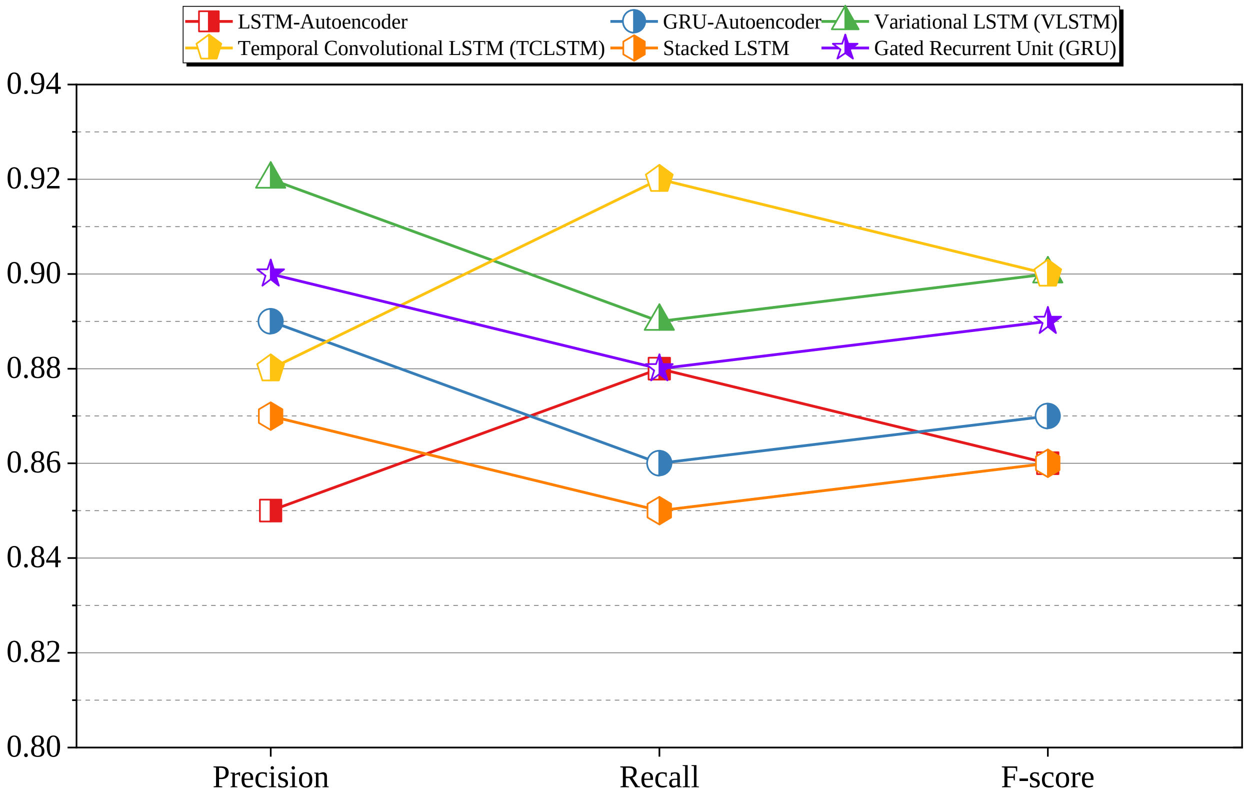


Fig. 2. Result of RNN-based methods.

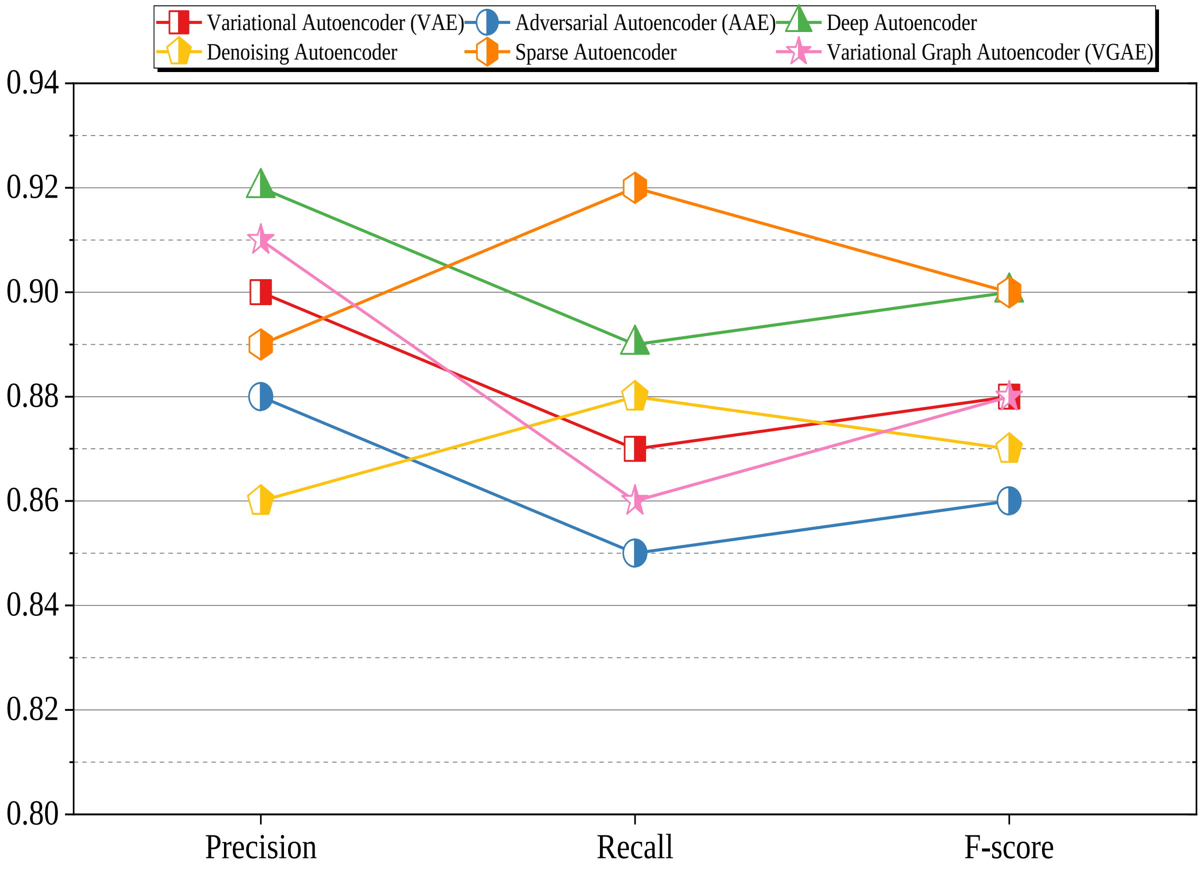


Fig. 3. Result of Autoencoders methods analysis.

D. Analysis of GANs-based Methods

This section presents a selection of GAN-based anomaly detection methods, along with their associated precision, recall, and F-score performance metrics. These methods include AnoGAN, Boundary-Seeking GAN (BGAN), Adversarial Variational Bayes (AVB), DualGAN, Energy-based GAN (EBGAN), and Generative Moment Matching Networks (GMMN).

As shown in Fig. 4, upon analyzing the table, it is clear that the performance of these GAN-based methods varies across different evaluation metrics. When considering precision, AVB stands out with a value of 0.90, indicating its ability to precisely identify anomalies while minimizing false positives compared to the other methods in the table. In terms of recall, DualGAN demonstrates the highest value at 0.89, suggesting its superior capability to capture a larger proportion of actual anomalies present in the data. Analyzing the F-scores, which provide a combined measure of precision and recall, AVB also outperforms other methods with an F-score of 0.88. This implies that AVB achieves a better balance between accurately identifying anomalies and minimizing false alarms compared to the other GAN-based methods. The better performance of AVB can be attributed to its ability to leverage the advantages of both adversarial learning and variational inference. By incorporating a variational autoencoder framework into the GAN architecture, AVB is able to model the underlying data distribution more effectively, resulting in improved precision, recall, and overall F-score.

E. Analysis of GCNs-based Methods

We select a collection of recent GCNs-based anomaly detection methods, along with their precision, recall, and F-score performance metrics. The selected methods include Graph Convolutional Autoencoder, GraphSAGE, Graph Attention Network (GAT), Deep Graph Convolutional Network (DGCN), Graph Convolutional LSTM (GC-LSTM), and Graph Isomorphism Network (GIN).

As shown in Fig. 5, upon analyzing the result data, it is evident that the GCNs-based methods exhibit varying performance across different evaluation metrics. Notably, GAT stands out in terms of precision, achieving an impressive value of 0.92. This indicates its exceptional ability to accurately identify anomalies while minimizing false positives compared to other methods listed in the table.

In the aspect of recall, GIN surpasses the rest with a score of 0.93, demonstrating its superior capability to capture a larger proportion of actual anomalies present in the data. Moreover, when considering the F-scores that provide a comprehensive measure of both precision and recall, GIN emerges as the top performer with an F-score of 0.91. This indicates that GIN strikes a better balance between accurately identifying anomalies and minimizing false alarms compared to other GCNs-based methods. The outstanding performance of GIN can be attributed to its innovative utilization of graph isomorphism as a fundamental concept within its design. By leveraging graph isomorphism, GIN effectively captures the underlying structural similarities and relationships in the data, leading to improved precision, recall, and overall F-score. This highlights the significance of incorporating domain-specific knowledge and leveraging graph-based representations to increase anomaly detection performance in GCNs-based approaches.

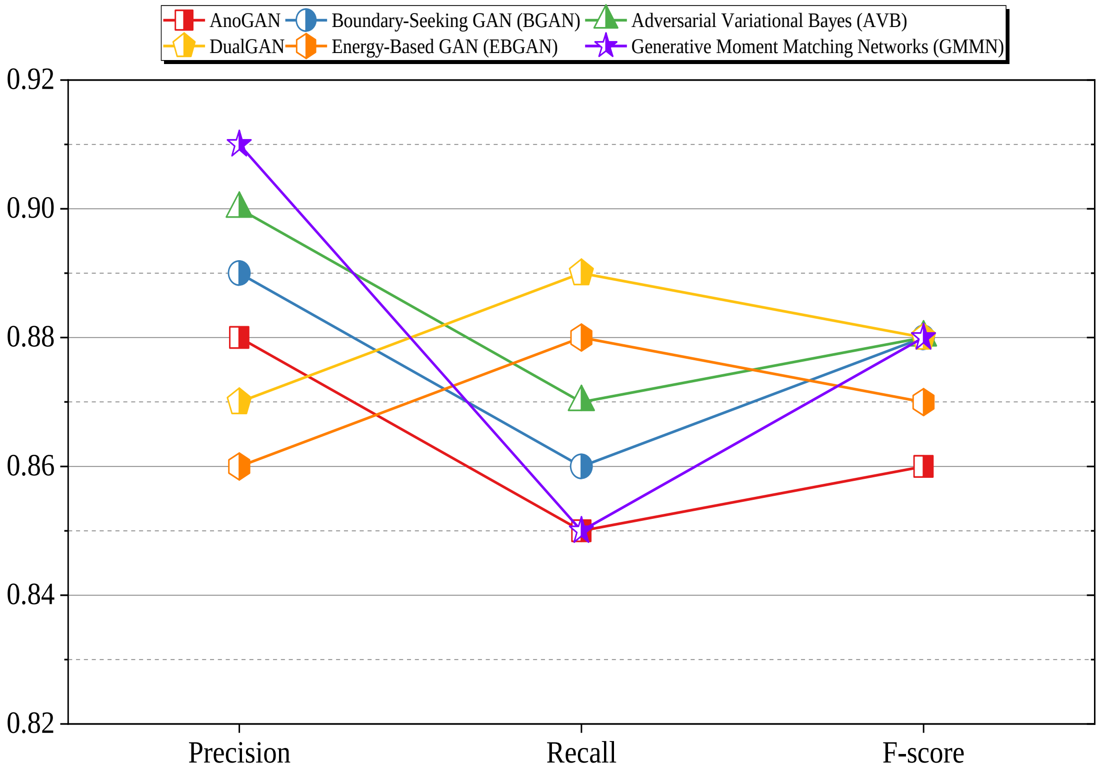


Fig. 4. Result of GANs-based methods.

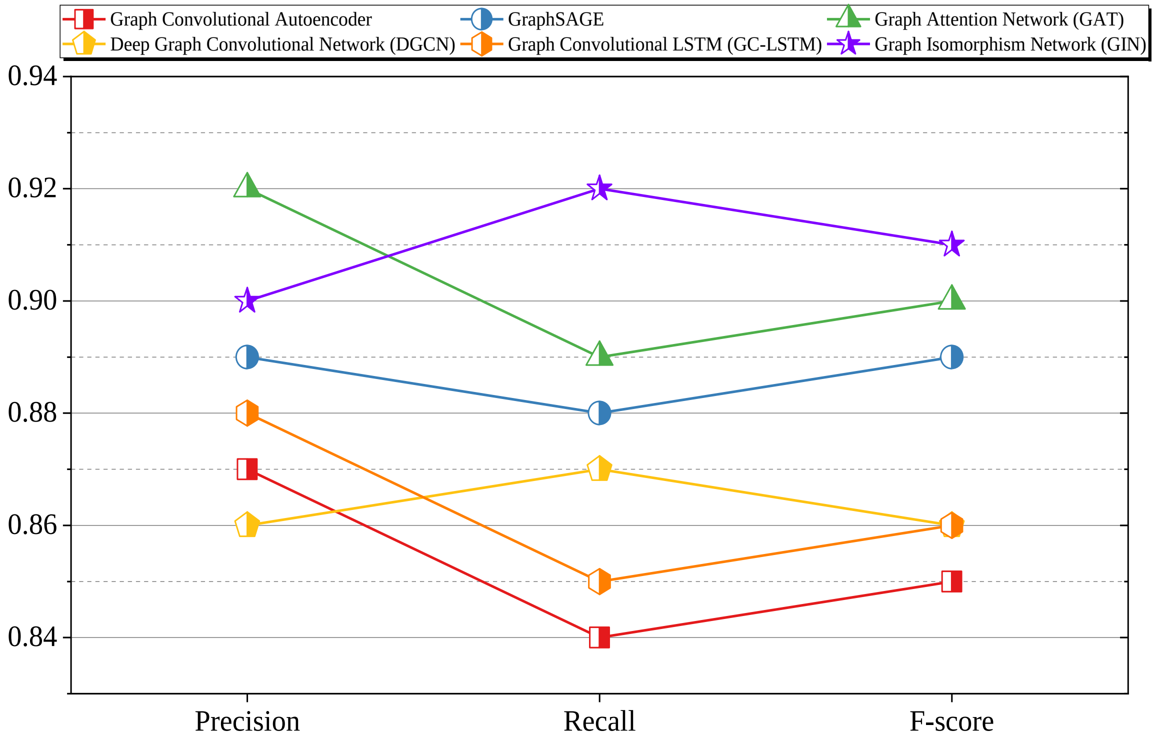


Fig. 5. Result of GCNs-based methods.

VI. Conclusion

In conclusion, this review paper sheds light on the significance of anomaly detection in IoT surveillance systems and the shift towards deep learning-based approaches to overwhelm the limitations of traditional methods. The justification for conducting a comprehensive review lies in the quest to identify the most accurate methods for anomaly detection. Through the collection of results and performance evaluations, this review paper provides a comprehensive analysis of existing deep learning techniques, bridging the research gaps in anomaly detection for IoT surveillance systems. By addressing these challenges and presenting a thorough examination of deep learning methods, this review paper paves the way for development of more efficient and effective anomaly detection solutions in the realm of IoT surveillance. For future work, one direction for further study is to concentrate on enhancing the interpretability of deep learning-based anomaly detection models for IoT surveillance systems. Developing techniques to explain the decisions and reasoning of these models can provide valuable insights into the detection process and build trust in their functionality. Exploring explainable AI methods, including attention mechanisms or feature visualization, can help in understanding the factors influencing anomaly detection and enable effective decision-making. Moreover, another important direction for future research is addressing the challenges posed by concept drift and dynamic environments in IoT surveillance systems. Anomaly detection models need to be adaptable and capable of continuously learning and updating their knowledge to accommodate changing patterns and emerging anomalies. Investigating techniques such as online learning, transfer learning, or adaptive models can facilitate the detection of evolving anomalies in real-time scenarios. Additionally, incorporating contextual information and temporal dependencies can enhance the models' ability to differentiate between normal variations and true anomalies in dynamic environments.

Acknowledgment

This work was supported by Science and technology research project of the Department of Education of Hebei province, research on crowd abnormal event detection and active directional warning system (No.QN2019184), and Research on teaching reform of program design curriculum based on “Four-in-one” in the Ministry of Education's industry-university Cooperation and education project in 2022(No.220606684213250).

References

[1] M. Islam, A. S. Dukyil, S. Alyahya, and S. Habib, "An IoT Enable Anomaly Detection System for Smart City Surveillance," Sensors, vol. 23, no. 4, p. 2358, 2023.

[2] F. T. Al-Dhief et al., "A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms," IEEE Access, vol. 8, pp. 64514-64533, 2020.

[3] O. Elharrouss, N. Almaadeed, and S. Al-Maadeed, "A review of video surveillance systems," Journal of Visual Communication and Image Representation, vol. 77, p. 103116, 2021.

[4] D. R. Patrikar and M. R. Parate, "Anomaly detection using edge computing in video surveillance system," International Journal of Multimedia Information Retrieval, vol. 11, no. 2, pp. 85-110, 2022.

[5] S. Jha, C. Seo, E. Yang, and G. P. Joshi, "Real time object detection and trackingsystem for video surveillance system," Multimedia Tools and Applications, vol. 80, pp. 3981-3996, 2021.

[6] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: recent updates, datasets, challenges, and applications," Artificial Intelligence Review, vol. 54, pp. 2259-2322, 2021.

[7] D. Chaudhary, S. Kumar, and V. S. Dhaka, "Video based human crowd analysis using machine learning: a survey," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 10, no. 2, pp. 113-131, 2022.

[8] A. Aghamohammadi, M. C. Ang, E. A. Sundararajan, K. W. Ng, M. Mogharrebi, and S. Y. Banihashem, "Correction: A parallel spatiotemporal saliency and discriminative online learning method for visual target tracking in aerial videos," Plos one, vol. 13, no. 3, p. e0195418, 2018.

[9] B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, and M. Khassanova, "State-of-the-art violence detection techniques in video surveillance security systems: a systematic review," PeerJ Computer Science, vol. 8, p. e920, 2022.

[10] L. Malphedwar and T. Rajesh, "Video based Anomaly Detection Utilizing the Crow Search Algorithm-based Deep RNN," Mathematical Statistician and Engineering Applications, vol. 71, no. 4, pp. 10-23, 2022.

[11] Z. Zhang, "Detecting Anomaly Event in Video Based on Generative Adversarial Network," Computational Intelligence and Neuroscience, vol. 2022, 2022.

[12] X. Ma et al., "A comprehensive survey on graph anomaly detection with deep learning," IEEE Transactions on Knowledge and Data Engineering, 2021.

[13] I. Ullah and Q. H. Mahmoud, "Design and development of a deep learning-based model for anomaly detection in IoT networks," IEEE Access, vol. 9, pp. 103906-103926, 2021.

[14] A. Aboah, "A vision-based system for traffic anomaly detection using deep learning and decision trees," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4207-4212.

[15] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," IEEE access, vol. 7, pp. 36322-36333, 2019.

[16] A. Aghamohammadi et al., "A deep learning model for ergonomics risk assessment and sports and health monitoring in self-occluded images," Signal, Image and Video Processing, pp. 1-13, 2023.

[17] J. Ren, F. Xia, Y. Liu, and I. Lee, "Deep video anomaly detection: Opportunities and challenges," in 2021 international conference on data mining workshops (ICDMW), 2021: IEEE, pp. 959-966.

[18] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," ACM computing surveys (CSUR), vol. 54, no. 2, pp. 1-38, 2021.

[19] R. Leyva, V. Sanchez, and C.-T. Li, "Video anomaly detection with compact feature sets for online performance," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3463-3478, 2017.

[20] R. J. Franklin and V. Dabbagol, "Anomaly detection in videos for video surveillance applications using neural networks," in 2020 Fourth International Conference on Inventive Systems and Control (ICISC), 2020: IEEE, pp. 632-637.

[21] R. Nayak, U. C. Pati, and S. K. Das, "A comprehensive review on deep learning-based methods for video anomaly detection," Image and Vision Computing, vol. 106, p. 104078, 2021.

[22] S. Chandrakala, K. Deepak, and G. Revathy, "Anomaly detection in surveillance videos: a thematic taxonomy of deep models, review and performance analysis," Artificial Intelligence Review, vol. 56, no. 4, pp. 3319-3368, 2023.

[23] D. Kwon, K. Natarajan, S. C. Suh, H. Kim, and J. Kim, "An empirical study on network anomaly detection using convolutional neural networks," in 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), 2018: IEEE, pp. 1595-1598.

[24] Z. Tang, Z. Chen, Y. Bao, and H. Li, "Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring," Structural Control and Health Monitoring, vol. 26, no. 1, p. e2296, 2019.

[25] H. Mu, R. Sun, M. Wang, and Z. Chen, "Spatio-temporal graph-based CNNs for anomaly detection in weakly-labeled videos," Information Processing & Management, vol. 59, no. 4, p. 102983, 2022.

[26] Y. Chang et al., "Video anomaly detection with spatio-temporal dissociation," Pattern Recognition, vol. 122, p. 108213, 2022.

[27] D. Koshti, S. Kamoji, N. Kalnad, S. Sreekumar, and S. Bhujbal, "Video anomaly detection using inflated 3d convolution network," in 2020 International Conference on Inventive Computation Technologies (ICICT), 2020: IEEE, pp. 729-733.

[28] R. Maqsood, U. I. Bajwa, G. Saleem, R. H. Raza, and M. W. Anwar, "Anomaly recognition from surveillance videos using 3D convolution neural network," Multimedia Tools and Applications, vol. 80, no. 12, pp. 18693-18716, 2021.

[29] M. Murugesan and S. Thilagamani, "Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network," Microprocessors and Microsystems, vol. 79, p. 103303, 2020.

[30] W. Luo et al., "Video anomaly detection with sparse coding inspired deep neural networks," IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 3, pp. 1070-1084, 2019.

[31] L. Bontemps, V. L. Cao, J. McDermott, and N.-A. Le-Khac, "Collective anomaly detection based on long short-term memory recurrent neural networks," in Future Data and Security Engineering: Third International Conference, FDSE 2016, Can Tho City, Vietnam, November 23-25, 2016, Proceedings 3, 2016: Springer, pp. 141-152.

[32] J. R. Medel and A. Savakis, "Anomaly detection in video using predictive convolutional long short-term memory networks," arXiv preprint arXiv:1612.00390, 2016.

[33] H. Fanta, Z. Shao, and L. Ma, "SiTGRU: single-tunnelled gated recurrent unit for abnormality detection," Information Sciences, vol. 524, pp. 15-32, 2020.

[34] P. Zhang and Y. Lu, "Research on Anomaly Detection of Surveillance Video Based on Branch-Fusion Net and CSAM," Sensors, vol. 23, no. 3, p. 1385, 2023.

[35] A. Ravi and F. Karray, "Exploring Convolutional Recurrent architectures for anomaly detection in videos: a comparative study," Discover Artificial Intelligence, vol. 1, pp. 1-16, 2021.

[36] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, and S. W. Baik, "An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos," Sensors, vol. 21, no. 8, p. 2811, 2021.

[37] Y. Chang, Z. Tu, W. Xie, and J. Yuan, "Clustering driven deep autoencoder for video anomaly detection," in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16, 2020: Springer, pp. 329-345.

[38] N. Li, F. Chang, and C. Liu, "Spatial-temporal cascade autoencoder for video anomaly detection in crowded scenes," IEEE Transactions on Multimedia, vol. 23, pp. 203-215, 2020.

[39] L. Wang, H. Tan, F. Zhou, W. Zuo, and P. Sun, "Unsupervised anomaly video detection via a double-flow convlstm variational autoencoder," IEEE Access, vol. 10, pp. 44278-44289, 2022.

[40] S. D. Bansod and A. V. Nandedkar, "Anomalous event detection and localization using stacked autoencoder," in Computer Vision and Image Processing: 4th International Conference, CVIP 2019, Jaipur, India, September 27–29, 2019, Revised Selected Papers, Part II 4, 2020: Springer, pp. 117-129.

[41] M. Ribeiro, M. Gutoski, A. E. Lazzaretti, and H. S. Lopes, "One-class classification in images and videos using a convolutional autoencoder with compact embedding," IEEE Access, vol. 8, pp. 86520-86535, 2020.

[42] C. Huang et al., "Self-supervised attentive generative adversarial networks for video anomaly detection," IEEE Transactions on Neural Networks and Learning Systems, 2022.

[43] X. Feng, D. Song, Y. Chen, Z. Chen, J. Ni, and H. Chen, "Convolutional transformer based dual discriminator generative adversarial networks for video anomaly detection," in Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 5546-5554.

[44] J. Montenegro and Y. Chung, "Semi-supervised generative adversarial networks for anomaly detection," in SHS Web of Conferences, 2022, vol. 132: EDP Sciences.

[45] D. Li, X. Nie, X. Li, Y. Zhang, and Y. Yin, "Context-related video anomaly detection via generative adversarial network," Pattern Recognition Letters, vol. 156, pp. 183-189, 2022.

[46] M. A. Contreras-Cruz, F. E. Correa-Tome, R. Lopez-Padilla, and J.-P. Ramirez-Paredes, "Generative Adversarial Networks for anomaly

detection in aerial images," Computers and Electrical Engineering, vol. 106, p. 108470, 2023.

[47] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," Medical image analysis, vol. 54, pp. 30-44, 2019.

[48] M. O. Kaplan and S. E. Alptekin, "An improved BiGAN based approach for anomaly detection," Procedia Computer Science, vol. 176, pp. 185-194, 2020.

[49] M. Ye, X. Peng, W. Gan, W. Wu, and Y. Qiao, "Anopcn: Video anomaly detection via deep predictive coding network," in Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1805-1813.

[50] W. Luo, W. Liu, and S. Gao, "Graph convolutional neural network for skeleton-based video abnormal behavior detection," in Generalization With Deep Learning: For Improvement On Sensing Capability: World Scientific, 2021, pp. 139-155.

A Graph-Cut Guided ROI Segmentation Algorithm with Lightweight Deep Learning Framework for Cervical Cancer Classification

Shiny T L[1], Kumar Parasuraman[2]

Research Scholar, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University,
Abishekapatti, Tirunelveli – 627 012, Tamil Nadu, India[1]
Associate Professor, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University,
Abishekapatti, Tirunelveli – 627 012, Tamilnadu, India[2]

***Abstract*—Cervical cancer classification has witnessed numerous advancements through deep learning methods; however, existing approaches often rely on multiple models for segmentation and classification, leading to heightened computational demands and prolonged training times. In this research, a lightweight deep learning framework for cervical cancer classification is presented. The framework comprises three primary components: a Graph-Cut Guided Region of Interest (ROI) segmentation algorithm, a streamlined DenseNet architecture, and a Multi-Class Logistic Regression classifier. The Graph-Cut Guided ROI segmentation algorithm is used to accurately isolate nuclei regions within multicellular Pap smear images. This is a lightweight algorithm that is able to achieve high segmentation accuracy with minimal computational overhead. The streamlined DenseNet architecture is used to efficiently extract salient features from the segmented images. This architecture is specifically designed to reduce feature redundancy and eliminate incongruous feature maps. The Multi-Class Logistic Regression classifier is used to classify the segmented images into different cell types and stages of cervical cancer. Experimental results show the proposed method is able to achieve high classification accuracy with minimal training time. The framework was trained and evaluated on a dataset of 963 Pap smear images. The proposed framework achieved a 98% cell type classification accuracy in precision, recall, and F1-score for classifying multi-cell Pap smear images. The training loss was also very low. The average training time was 21 minutes for different sets of training images, and the average testing time was 0.50 seconds for different sizes of testing images, which is much lower than the existing methods.**

Keywords*—Cervical cancer classification; deep learning; lightweight deep learning framework; graph-cut guided ROI segmentation algorithm; nuclei region isolation

I. INTRODUCTION

Cervical cancer is the fourth most common cancer among women worldwide, with a substantial impact on public health [1]. The Pap smear, a widely used screening method, involves the microscopic examination of cervical cells to detect abnormalities and enable early intervention [2], [3]. Automated classification of these cells can significantly expedite the diagnosis process and facilitate timely medical intervention. In recent times, deep learning techniques have emerged as powerful tools for medical image analysis, with numerous methodologies proposed for cervical cancer classification While deep learning models have demonstrated impressive performance in various medical imaging tasks, they often require extensive computational resources, intricate model architectures, and prolonged training times [4], [5], [6], [7]. Many existing approaches to cervical cancer classification utilize separate models for segmentation and classification [8], [9], [16], [20]. The paradigm of utilizing separate models for segmentation and classification exacerbates the computational demands of the overall system. The need for multiple models not only increases resource requirements but also extends the time required for model training and inference. This can hinder the integration of automated cervical cancer classification systems into clinical practice, where prompt diagnosis is crucial for effective treatment planning. Additionally, some existing approaches employ feature fusion methods to combine the outputs of separate models [10], [11], [14], [18]. While these methods attempt to leverage the strengths of different models, they can introduce further complexity and potential sources of error. Integrating outputs from multiple models requires careful consideration of weightings and fusion strategies, and may not always result in optimal performance. Moreover, many existing methods for segmenting and classifying cervical cancer cells in pap smear images are only tested with single cell images [12], [13], [14]. However, in reality, Pap smear images often contain multiple cells, which can make it more difficult to accurately segment and classify the cells.

This research endeavors to address these limitations by introducing an innovative methodology that enhances computational efficiency while maintaining robust classification accuracy. To mitigate the computational burden and accelerate the time-consuming segmentation process, a lightweight Graph-Cut Guided ROI segmentation algorithm is introduced this research. This algorithm utilizes graph-cut techniques to effectively segment nuclei regions from multicell Pap smear images. By incorporating graph-cut guidance, this proposed approach significantly improves segmentation accuracy while maintaining computational efficiency. This innovation not only expedites the preprocessing step but also ensures that accurate nuclei regions are extracted for subsequent feature extraction. To extract the meaningful and relevant features from the segmented Pap smear images, this research employs a lightweight DenseNet architecture. DenseNet is renowned for its ability to alleviate the vanishing gradient problem and promote feature reuse across layers. The modified DenseNet architecture is tailored to the characteristics of cervical cell images, enabling efficient feature extraction while minimizing feature redundancy. The architecture's capacity to eliminate uncorrelated feature maps further enhances computational efficiency without compromising classification accuracy. After feature extraction, the next crucial step is accurate classification. A Multi-Class Logistic Regression model is adopted for this purpose. This model is well-suited for scenarios with multiple classes, such as different stages of cervical cancer. Logistic Regression is known for its simplicity and interpretability, while still delivering commendable classification performance. Integrating this classifier into the proposed architecture preserves computational efficiency without compromising the accuracy of the classification process. The collaboration between these key components results in an integrated methodology that offers several notable advantages:

1) *Computational efficiency:* By integrating lightweight Graph-Cut Guided ROI segmentation and an efficient DenseNet architecture significantly reduces computational demands.

2) *Streamlined pipeline:* The integration of segmentation, feature extraction, and classification into a cohesive pipeline eliminates the need for multiple models, simplifying the overall process and reducing model training time.

3) *Enhanced accuracy*: The precision of the segmentation process and the efficacy of the feature extraction mechanism contribute to improved classification accuracy. Accurate segmentation ensures that only relevant regions are considered for feature extraction, while the DenseNet architecture captures essential image characteristics.

Following section of the article reviews recently developed convolutional neural network (CNN) methods for cervical cancer detection. Section II outlines the literature review. Section III presents the proposed deep learning-based framework for cervical cancer classification. Section IV discusses the experimental analysis and results. Finally, Section V the paper concludes with a summary of the key findings and implications.

II. Literature Review

The literature review reveals a range of approaches aimed at improving cervical cancer classification using fusion and multi-model CNNs. These methods address the challenge of accurate and efficient classification while considering the complex nature of medical image analysis. However, each approach comes with its own strengths and limitations.

Md Mamunur Rahaman et al. [14]: This research proposes the DeepCervix approach, leveraging a hybrid deep feature fusion technique. By integrating XceptionNet, ResNet50, VGG19, and VG16 models, this method addresses cervical cell classification. The fusion process occurs after preprocessing, followed by fine-tuning specific layers. However, challenges might arise in effectively combining features from diverse models, potentially affecting performance and interpretability. Ponnusamy Sukumar et al. [15]: The study introduces an automated framework for cervical cancer classification in Pap smear images. Employing watershed segmentation for cell nucleus segmentation and a feature extraction strategy involving Local Binary Pattern, texture, GLCM, laws texture, and histogram features, the approach utilizes principal component analysis for classification. The reliance on watershed segmentation could lead to issues due to local minima and noise sensitivity. Zaid Alyafeai et al. [16]: A pipeline-based architecture is suggested, featuring two pre-trained deep models for cervix region identification and tumor classification. Through modules like ROI detection, pre-processing with data augmentation, and lightweight CNN-based feature extraction, the method strives to address segmentation and classification. However, this approach's complexity, dependence on specific pre-trained models, and potential challenges in integration. Yuexiang Li et al. [17]: This research presents a model with a key-frame feature encoding network and a feature fusion network. By encoding pre- and post-acetic acid test images using ResNet-101 and employing an interpretable graph convolution network (E-GCN) for classification, the method focuses on feature fusion. The interpretability of the graph convolution network might present difficulties in a clinical setting.

Long D. Nguyen et al. [18]: An ensemble method involving feature concatenation and deep CNNs is introduced for cervical cancer classification. Combining features from InceptionResNet-v2, Inception-v3, and ResNet152, the method seeks to enhance classification. However, the substantial ensemble of CNNs might lead to resource-intensive computations and complexity. Srishti Gautam et al. [19]: The proposed deep learning-based approach addresses detection, segmentation, and classification of nuclei in single cell Pap smear images. While the focus on feature-based segmentation and AlexNet classification through transfer learning is promising, the method's applicability to diverse nucleus types should be evaluated. Deepa. K et al. [20]: This deep learning architecture emphasizes the classification of cervical cancer using a combination of segmentation and feature extraction techniques. The use of Mask Region Based Convolutional Neural Network for segmentation and subsequent feature extraction from segmented images with GLCM, Contourlet, and Gabor filters contributes to classification using VGG networks. The intricacies of combining multiple feature extraction techniques merit investigation. Swati Shinde et al. – DeepyCyto [21]: DeepyCyto proposes a hybrid methodology utilizing two workflows for cervical cancer classification. Integrating feature fusion vectors from pre-trained models, this approach applies machine learning ensemble and artificial neural network methods. Challenges in computational cost and adaptability to overlapping images could limit its feasibility.

III. Proposed method

The proposed research is a deep learning-based framework for cervical cancer classification. The framework consists of five modules: dataset, contrast enhancement, graph-cut guided ROI segmentation, image augmentation, and feature extraction using lightweight DenseNet. The dataset consists of 963 pap smear images with four distinct classes: HSIL, LSIL, NIM, and

SCC. Contrast enhancement is applied to improve nuclei visibility and aid feature extraction.

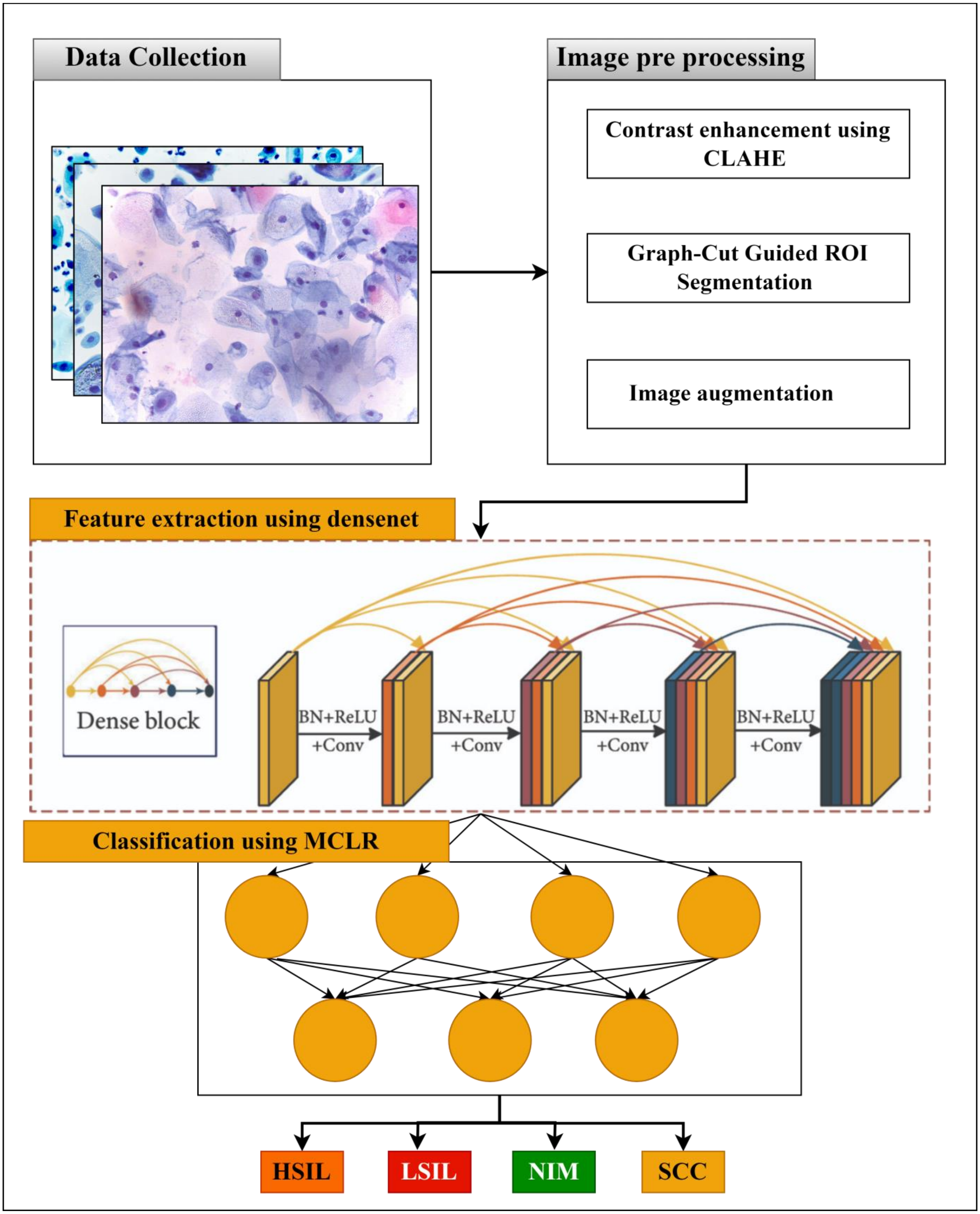


Fig. 1. Overall architecture of proposed cervical cancer classification system.

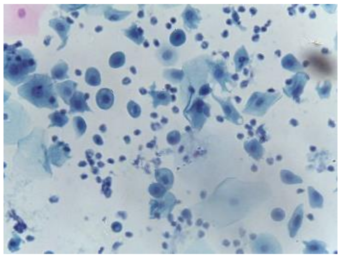
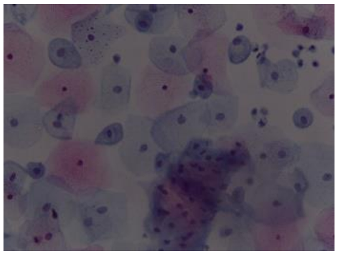
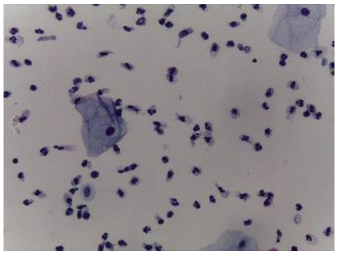
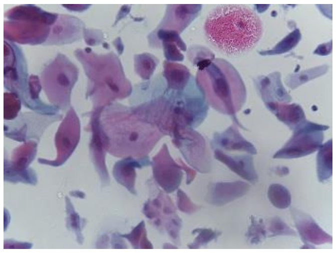
Graph-cut guided ROI segmentation is used to segment nuclei regions. Image augmentation is applied to enhance dataset diversity and generalize the deep learning model. Feature extraction using lightweight DenseNet is used to extract informative features from the segmented images. Finally, a Multi-Class Logistic Regression classifier is trained using the extracted features to classify the images into the four classes. Fig. 1 shows the overall architecture of proposed cervical cancer classification system.

A. Dataset Details

The dataset utilized in this research is named "Liquid based cytology pap smear images" [22].

This dataset was sourced from https://data.mendeley.com /datasets/zddtpgzv63/4, and it consists of a collection of 963 Pap smear images. These images were obtained using 40x magnification through the Leica ICC50 HD microscope. The dataset encompasses four distinct classes, each representing specific diagnostic categories. These classes are as follows: High Squamous Intraepithelial Lesion (HSIL), Low Squamous Intraepithelial Lesion (LSIL), Negative for Intraepithelial Malignancy (NIM), and Squamous Cell Carcinoma (SCC). These classes reflect different pathological conditions of cervical cells, enabling the classification of various stages of cervical cancer and cell types. Table I explains the different pap smear in the dataset.

TABLE I. DATASET OF PAP SMEAR IMAGES WITH CLASS NAMES AND NUMBER OF IMAGES

Image name	Image	Image details	Class name	Number of images
High Squamous Intraepithelial Lesion		HSIL refers to significant and advanced abnormal changes in the squamous epithelial cells of the cervix. These changes are considered more severe than low-grade lesions. HSIL is characterized by cells that appear markedly abnormal, with a higher likelihood of progressing to cervical cancer if left untreated [23].	HSIL	253
Low Squamous Intraepithelial Lesion		LSIL indicates milder abnormal cellular changes in the squamous epithelial cells of the cervix. These changes are not as severe as HSIL. While LSIL is generally considered less likely to progress to cancer than HSIL, it's still important to monitor and possibly treat these cases to prevent progression.	LSIL	198
Negative for Intraepithelial Malignancy.		This classification indicates that the cervical cell sample appears normal and lacks any signs of intraepithelial malignancy.	NIM	252
Squamous Cell Carcinoma.		Squamous cell carcinoma is a type of cervical cancer that originates from the squamous epithelial cells of the cervix. This classification indicates the presence of cancerous cells within the cervical sample. It's a more advanced stage of abnormality compared to HSIL, as it indicates the presence of actively dividing and invasive cancer cells.	SCC	260

B. Contrast Enhancement

Cervical cell images captured in Pap smears can have variations in intensity due to different acquisition conditions, staining procedures, and sample preparation techniques. These variations can make it challenging to differentiate between important structures like nuclei and background. Accurate nuclei detection relies on clear distinctions between nuclei and the surrounding background. Contrast enhancement enhances the visibility of nuclei, making them easier to detect. The streamlined DenseNet architecture relies on extracting meaningful features. Enhanced contrast helps the network identify subtle textures and structures, leading to more informative features. An accurate Multi-Class Logistic Regression classifier requires well-defined features. Enhanced contrast can make the differences between different cell types and stages of cervical cancer more pronounced, leading to improved classification accuracy. In this research, Contrast-Limited Adaptive Histogram Equalization (CLAHE) used to normalize the contrast of input pap smear image [24]. CLAHE enhances the contrast of an image while preventing excessive amplification of noise. It works by dividing the image intsmaller regions, applying histogram equalization to each region, and then limiting the contrast amplification. The process flow of contrast enhancement is explained in Algorithm 1.

Algorithm 1. CLAHE.

Step 1: Divide the Image into Tiles:

Divide the image into tiles of size (w, h).

Step 2: Compute Histograms:

For each tile, calculate the histogram for each colour channel (R, G, and B).

Step 3: Histogram Equalization:

Apply histogram equalization to each tile's histogram to redistribute pixel intensity values and enhance local contrast.

The equation for histogram equalization is:

$$\boldsymbol{f(x) = L - 1 - (L - 1) \times \frac{cdf(x)}{sum(cdf)}} \qquad [1]$$

where:

- **f(x) is the output intensity value for input intensity value x.**
- **L is the number of intensity levels in the image.**
- **cdf(x) is the cumulative distribution function of the input histogram.**
- **sum(cdf) is the sum of the values in the input histogram.**

Step 4: Clip Histograms:

To avoid over-amplification of noise, clip the histogram bins to a specified threshold.

The equation for clipping the histogram is:

$$if\ (x > threshold)\, \{\ x = threshold; \}$$

where:

- x is the current intensity value
- threshold is the clipping threshold

Step 5: Interpolation:

Smooth out intensity transitions between tiles by using interpolation techniques to maintain a coherent image appearance.

One common interpolation technique is bilinear interpolation, which is given by the following equation:

$$\boldsymbol{y = (x - x0) \times (y1 - y2) + (x1 - x) \times \frac{(y0 - y2)}{(x1 - x0)}} \qquad [2]$$

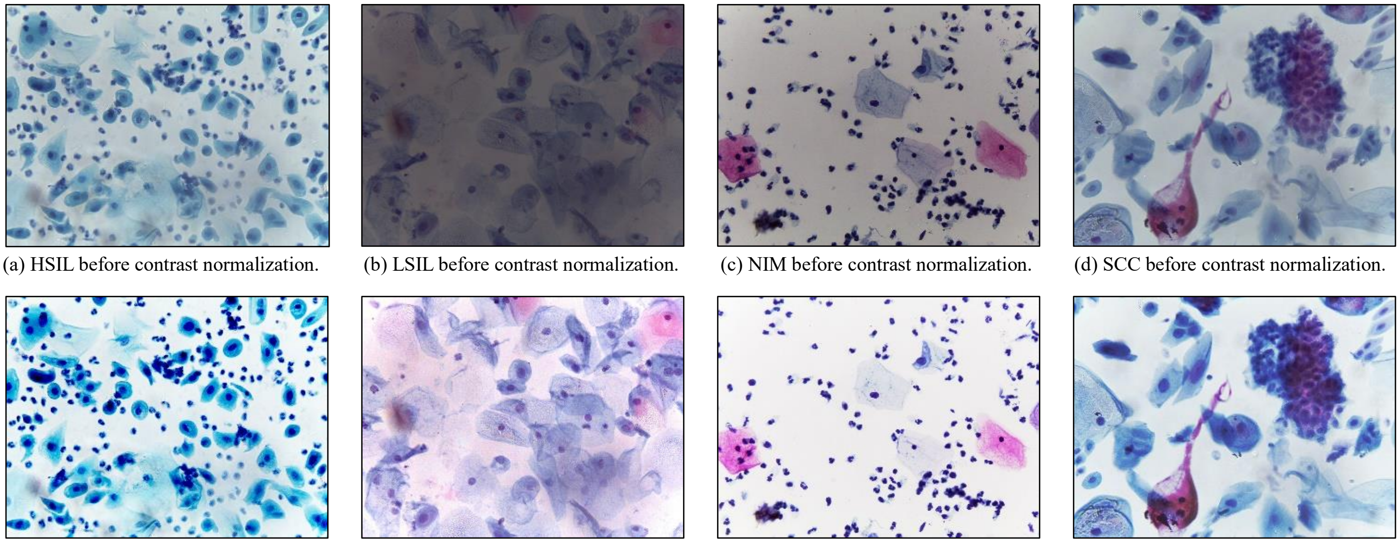
where:

- y is the output intensity value
- x is the input intensity value
- x0 and x1 are the x-coordinates of the two neighbouring pixels
- y0 and y1 are the y-coordinates of the two neighbouring pixels

Step 6: Reconstruct the Image:

Merge the enhanced tiles to reconstruct the full RGB image.

Fig. 2 illustrates the side-by-side comparison of Pap smear images before and after contrast normalization across different classes. The clear visual evidence showcases the remarkable impact of contrast enhancement using the CLAHE method. This enhancement process noticeably normalizes the contrast of the initial Pap smear images, resulting in enhanced image quality and improved visual clarity.



(a) HSIL before contrast normalization. (b) LSIL before contrast normalization. (c) NIM before contrast normalization. (d) SCC before contrast normalization.

(e) HSIL after contrast normalization. (f) LSIL after contrast normalization. (g) NIM after contrast normalization. (h) SCC after contrast normalization.

Fig. 2. Comparison of pap smear images before and after contrast normalization across different classes.

C. Proposed Graph-Cut Guided ROI Segmentation

The proposed graph-cut guided ROI segmentation algorithm is a way to divide an image into different regions of interest (ROIs). It works by using the principles of graph theory and energy minimization. By constructing a graph representation of the image, where pixels are nodes and their relationships are edges, the algorithm seeks to partition the image into foreground and background regions. Leveraging the concept of energy, the algorithm minimizes a cost function that encapsulates both local and global image characteristics. This approach yields detailed segmentation results, accurately defining nuclei regions from the complex background of multicell Pap smear images. The following steps outline the process of this segmentation method:

Algorithm. 2 graph-cut guided ROI segmentation

Step 1: Preprocessing

1.1 Input Image Preparation: The algorithm takes an input image I consisting of multicell Pap smear data. The image is represented as a grid of pixels, where each pixel is a node in the graph.

Step 2: Nuclei Detection and Seed Generation

2.1 Nuclei Detection: Nuclei are detected within the image I using method, S , that produces a binary mask M .

$$M = S(I) \quad [3]$$

2.2 Seed Point Generation: For each nucleus, a seed point $p\,i$ is generated. These seed points serve as initial markers for the graph-cut segmentation process.

$$p_i = (x_i, y_i), i = 1, 2, \dots, N \quad [4]$$

Step 3: Graph Construction

3.1 Construction of Graph: A graph $G = (V , E)$ is constructed, where V is the set of nodes (pixels) and E is the set of edges. Each pixel $v\,i\,j$ is a node in the graph.

$$V = \{v_{ij}\}, i = 1, 2, \dots, H; j = 1, 2, \dots, W \quad [5]$$

3.2 Edge Weights: Edge weights $w\,i\,j$, are assigned based on pixel intensities and spatial relationships.

$$wij, kl = exp\left(-\frac{(Iij - Ikl)^2}{2\sigma I^2}\right) \cdot exp\left(-dij, \frac{kl^2}{2\sigma d^2}\right) \quad [6]$$

Step 4: Graph-Cut Segmentation

4.1 Energy Minimization: The energy E is minimized to find the optimal cut that separates the foreground and background regions. The energy function involves unary and pairwise terms.

$$E = \sum_i D_i(s_i) + \sum_{i,j} V_{ij}(s_i, s_j) \quad [7]$$

Where $D\,i$ is the data cost, $V\,i\,j$ is the pairwise cost, and $s\,i$ is the label of node $v\,i\,j$.

4.2 Seed Propagation: The segmented regions S obtained from the graph-cut process are refined using a seed propagation mechanism.

$$S = SeedPropagate(M, S) \quad [8]$$

Step 5: ROI Extraction

5.1 Region of Interest Extraction: The segmented nuclei regions R are extracted from the input image I using the refined segmentation S .

$$R = I \odot S \quad [9]$$

Step 6: Postprocessing

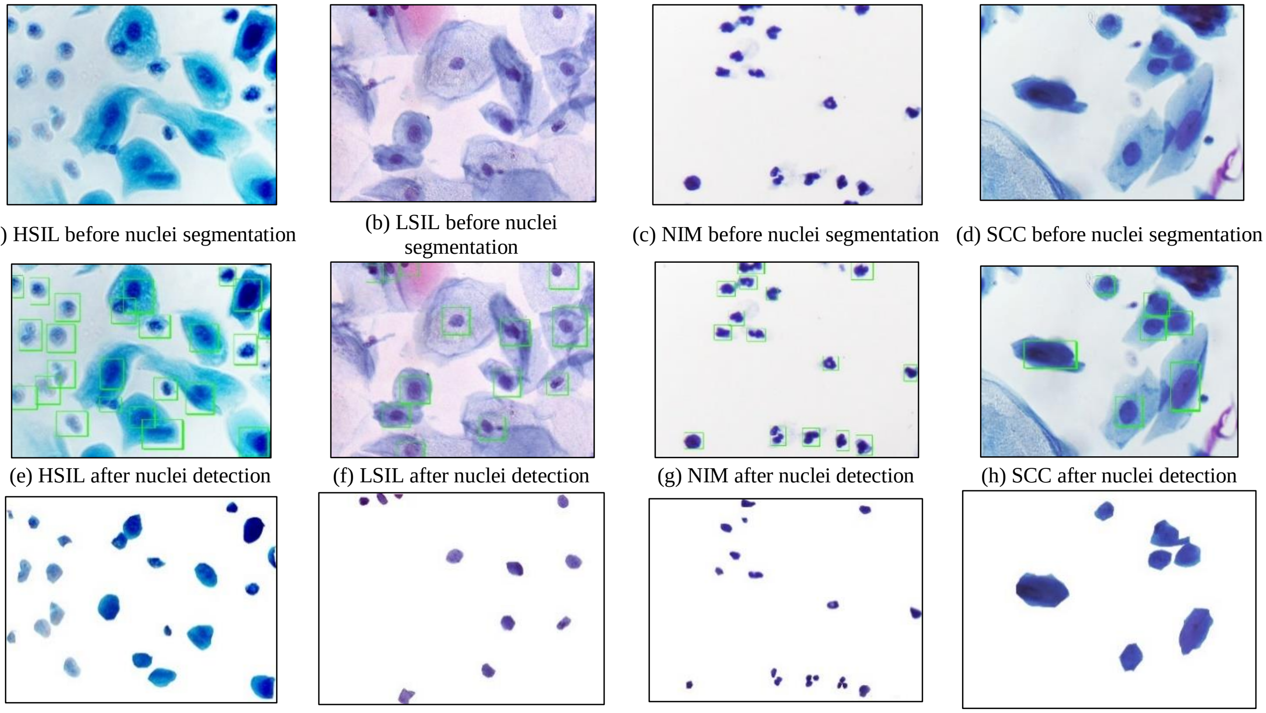
6.1 Noise Reduction: Postprocessing techniques such as morphological operations and noise reduction filters are applied to R to remove small artifacts.

$$R = MorphologicalOperations(R) \quad [10]$$

Step 7: Output Generation

7.1 Segmentation Results: The final segmented nuclei regions R are obtained as output, representing accurate identification of nuclei within the multicell Pap smear image.

The Fig. 3 shows the results of nuclei segmentation for four different cervical cancer grades: HSIL, LSIL, NIM, and SCC using proposed Graph-Cut Guided ROI Segmentation algorithm. The algorithm's effectiveness in accurately isolating nuclei regions within multicellular Pap smear images is demonstrated through these segmentation results.



(a) HSIL before nuclei segmentation

(b) LSIL before nuclei segmentation

(c) NIM before nuclei segmentation

(d) SCC before nuclei segmentation

(e) HSIL after nuclei detection

(f) LSIL after nuclei detection

(g) NIM after nuclei detection

(h) SCC after nuclei detection

(i) HSIL segmentation results

(j) LSIL segmentation results

(k) NIM segmentation results

(l) SCC segmentation results

Fig. 3. Nuclei segmentation results for different cervical cancer grades using proposed Graph-Cut Guided ROI Segmentation algorithm.

D. Image Augmentation

Image augmentation is a crucial technique employed to enhance the diversity of the dataset and improve the generalization ability of the deep learning model [25]. This research, a set of image augmentation techniques is applied to the original pap smear images before being used for training. The augmentation techniques include rotation, horizontal and vertical flipping, random zooming, and brightness adjustments. These transformations introduce variations that reflect real-world conditions and different microscope orientations, thereby minimizing the risk of overfitting and improving the model's ability to generalize to unseen data. Fig. 4 shows the different image augmentation techniques used in this research for enhanced Pap smear analysis. Data augmentation expanded the dataset from 963 to 5,778 images.

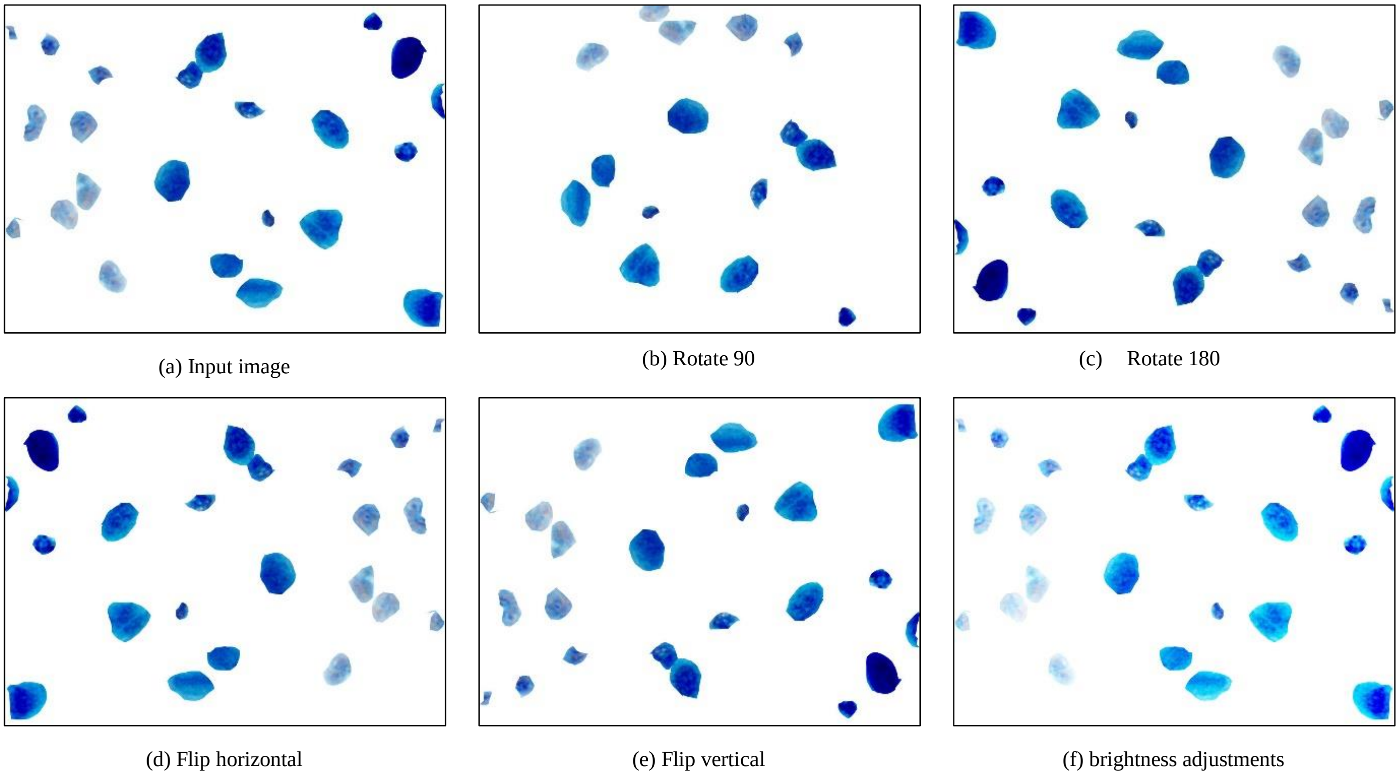

(a) Input image (b) Rotate 90 (c) Rotate 180
(d) Flip horizontal (e) Flip vertical (f) brightness adjustments

Fig. 4. Image augmentation techniques for enhanced Pap smear analysis.

E. Feature Extraction using Proposed Lightweight DenseNet

DenseNet is particularly well-suited for feature extraction from medical images due to its dense connectivity pattern [26]. Traditional convolutional architectures lose information as they progress through layers. DenseNet mitigates this by connecting each layer to every subsequent layer, ensuring that all features are directly accessible by subsequent layers [27]. This not only enhances feature propagation but also reduces the risk of information loss, making it ideal for capturing intricate structures within segmented medical images. The DenseNet architecture consists of dense blocks, each composed of multiple convolutional layers, followed by a transition layer. The key components are:

1) Dense block: Multiple convolutional layers are stacked together, and each layer receives the feature maps from all preceding layers. This dense connectivity encourages feature reuse and fosters information flow.

2) Transition layer: It follows each dense block and comprises convolutional and pooling layers. Transition layers down-sample feature maps, reducing computational load and channel dimensions, which aids in compactly representing salient features.

To cater to the characteristics of segmented Pap smear images, a modification is introduced to the traditional DenseNet architecture. The modification's primary goals are to reduce feature redundancy and enhance feature correlation, leading to more efficient and meaningful feature extraction. In this modified DenseNet architecture, feature fusion and channel attention mechanisms are integrated within each dense block. Feature fusion combines the outputs of different convolutional layers, ensuring complementary information is considered. Channel attention recalibrates feature maps, emphasizing important channels and reducing redundancy. Table II explain the architecture details of proposed DenseNet. Fig. 5 shows the visual representation of the proposed DenseNet. The initial convolutional layer processes the input, followed by multiple dense blocks, which enhance feature extraction capabilities. Transition layers between dense blocks down-sample the feature maps, reducing computational load. The introduced modification for feature fusion and channel attention would be integrated within the dense blocks.

Specifically, the feature fusion mechanism combines features from different convolutional layers within each dense block. The channel attention mechanism recalibrates feature maps, improving feature correlation and reducing redundancy. This architecture effectively exploits the dense connectivity pattern of DenseNet, encouraging efficient feature reuse and hierarchical representation of image structures. The modification further enhances the model's ability to extract meaningful features from segmented Pap smear images, contributing to accurate cervical cancer classification.

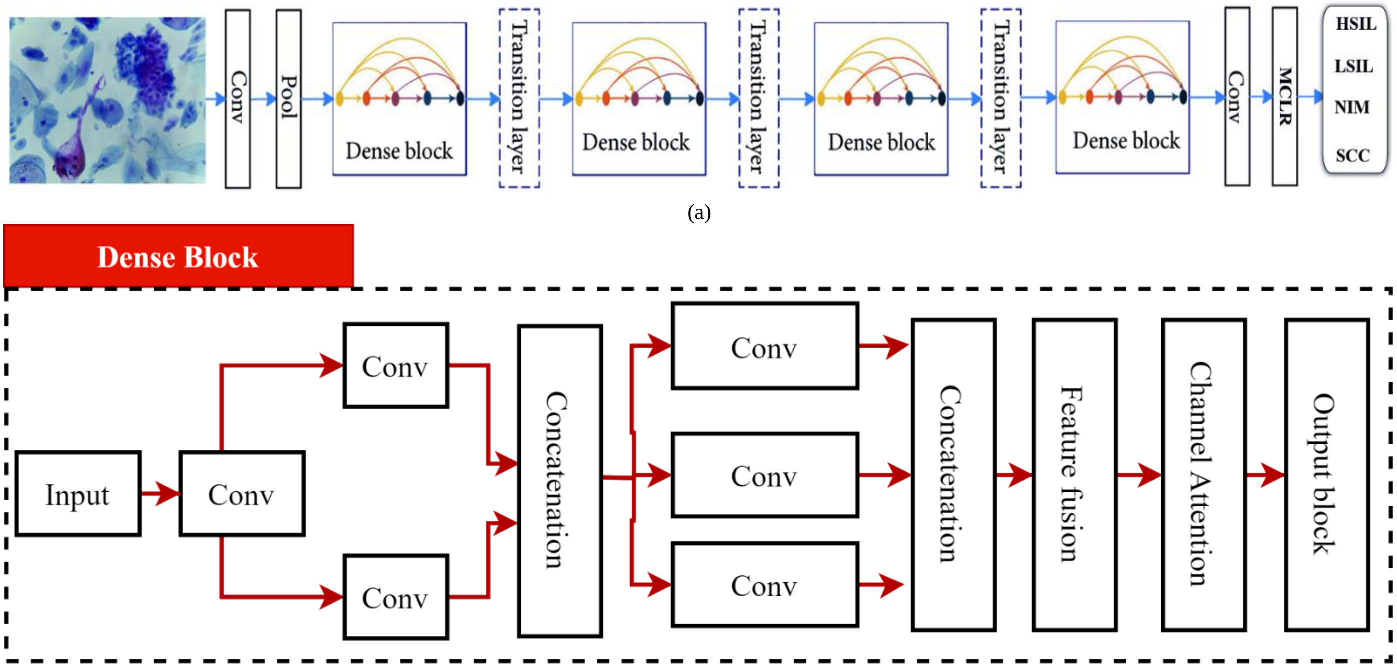
F. Classification

The classification module of the research involves training a classifier to differentiate between different cervical cancer grades based on the features extracted from the segmented and enhanced pap smear images. A Multi-Class Logistic Regression (MCLR) classifier is employed for this purpose, leveraging the informative features obtained through the modified DenseNet architecture. The MCLR classifier assigns a probability distribution over the classes, allowing for the identification of the most likely class for each input image. Given an input feature vector **x**, the MCLR model computes the probabilities for each class $\mathbf{C}i$ using the softmax function. The probability ($\mathbf{C}i \mid \mathbf{x}$) represents the likelihood that the input **x** belongs to class $\mathbf{C}i$. The class with the highest probability is predicted as the final output. The equation for the softmax function is:

$$(Ci|x) = \frac{e^{zi}}{\sum_{j=1}^{K} e^{zj}} \qquad (11)$$

TABLE II. ARCHITECTURE DETAILS OF PROPOSED DENSENET

Layer	Type	Configuration
Input	Input	224x224x3 (RGB with segmentation mask channel)
Preprocessing	Normalization	Normalize pixel values
Initial Convolution	Convolution	7x7 filter, 64 filters, stride = 2, padding = 'same'
	Batch Normalization	
	ReLU Activation	
	Max Pooling	3x3 pool, stride = 2
Dense Block	Batch Normalization	
	ReLU Activation	
	Convolution (1x1)	k filters (bottleneck), padding = 'same'
	Batch Normalization	
	ReLU Activation	
	Convolution (3x3)	k filters, padding = 'same'
Transition Layer	Batch Normalization	
	ReLU Activation	
	Convolution (1x1)	k filters (reducing channels), padding = 'same'
	Average Pooling	2x2 pool, stride = 2
Global Average Pooling	Global Average Pooling	Across spatial dimensions
Fully Connected Layers	Flatten	Flatten pooled feature maps



(a)

(b)

Fig. 5. Proposed DenseNet architecture (a) overall architecture (b) layer details of the proposed DenseNet block.

where: zi is the logit for class $\mathbf{C}i$ calculated as a linear combination of the feature vector $\mathbf{x}$ and the class-specific weights. K is the total number of classes. The predicted class for an input image is the one corresponding to the highest probability among all classes. The MCLR classifier is trained using the cross-entropy loss function, it measures the dissimilarity between the predicted probabilities and the actual class labels in the training data. The formula for the Cross-Entropy Loss is as follows:

$$H(y,p) = -sum_i y_i log(p_i) \tag{12}$$

where: y is the ground truth label (a vector of 1s and 0s, where 1 indicates the correct class and 0 indicates the incorrect class). p is the predicted probability distribution (a vector of probabilities, where each element represents the probability of the input being in the corresponding classes: HSIL, LSIL, NIM, and SCC). i is the index of a class and log is the natural logarithm.

IV. Results and Discussion

In this section, a comprehensive analysis of the results obtained from the proposed cervical cancer detection method is presented. This section is divided into three significant parts, each contributing to a deeper understanding of the effectiveness and efficiency of the proposed approach. Firstly, the accuracy of the proposed method is evaluated in comparison to existing deep learning models and recently published cervical cancer detection systems discussed in the literature review. Specifically, a comparison is made regarding the accuracy achieved by the method against popular deep learning architectures such as AlexNet, GoogleNet, VGGNet, ResNet, and YOLO. Additionally, the method's performance is assessed against recently developed systems like the DeepCervix approach by Md Mamunur Rahaman et al. [14], utilizing a hybrid deep feature fusion technique. Furthermore, analysis is conducted on the pipeline-based architecture proposed by Alyafeai et al. [16], featuring two pre-trained deep models for cervix region identification and tumor classification. The model presented by Yuexiang Li et al. [17], which incorporates a key-frame feature encoding network and a feature fusion network, is also considered. Lastly, the deep learning architecture by Deepa. K et al. [20], emphasizing cervical cancer classification through a combination of segmentation and feature extraction techniques using two CNN models, is examined. Following this, the training efficiency of the proposed method is evaluated. Lastly, a comparison is made regarding the computational efficiency of the proposed method with existing deep learning models and recently published cervical cancer detection systems. Table III presents the parameter settings for both conventional techniques and existing deep learning models.

A. System Details

The research was conducted using a well-configured system that incorporated artificial intelligence software and hardware components. The software components included MATLAB 2018, operating on the Windows 11 operating system. Regarding the hardware details, the system featured an Intel Core i7 processor. The system was also equipped with an NVIDIA GPU. Additionally, the system was equipped with 16GB of RAM.

B. Accuracy Analysis

Accuracy, precession, recall, and F1-measure are used to evaluate the accuracy of the proposed cervical cancer classification system. These are based on true positive cervical cancer classification (TP), true negative cervical cancer classification (TN), false positive cervical cancer classification (FP), and false negative cervical cancer classification (FN). Accuracy metrics are calculated using the following formula. The proposed cervical cancer classification system's accuracy was tested using four different types of cervical cancer cells: HSIL, LSIL, NIM, and SCC (see Section III (A)). The samples were all multicell Pap smear images.

$$Accuracy = \frac{(TP+TN)}{Total\ samples} \tag{13}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{14}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{15}$$

$$F1 - measure = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{16}$$

The proposed methodology's performance is evaluated against several existing deep learning models. The comparison is presented in Tables IV, V, VI, and VII.

TABLE III. Parameter Settings for Conventional Techniques and Existing Deep Learning Models

Parameter	Values
Learning Rate	0.001
Batch Size	64
Activation Function	ReLU
Dropout Rate	0.2
Data Augmentation	Yes
Optimizer	Adam
Weight Initialization	Random

TABLE IV. Classification Efficiency Comparison for HSIL Cells Classification with Deep Learning Models

Method	Accuracy	Precision	Recall	F1-Measure
AlexNet	91.25%	95.69%	91.9%	93.29%
GoogleNet	93.28%	96.70%	92.60%	94.60%
VGGNet	94.00%	97.20%	93.10%	95.10%
ResNet	94.60%	97.70%	95.60%	95.70%
YOLO	96.72%	97.80%	94.60%	95.70%
Proposed Method	98.15%	98.52%	97.33%	98.43%

TABLE V. CLASSIFICATION EFFICIENCY COMPARISON FOR LSIL CLASSIFICATION WITH DEEP LEARNING MODELS

Method	Accuracy	Precision	Recall	F1-Measure
AlexNet	91.72%	93.41%	94.91%	93.61%
GoogleNet	92.71%	94.91%	94.59%	94.78%
VGGNet	93.41%	95.40%	95.70%	95.20%
ResNet	94.99%	96.60%	95.51%	96.21%
YOLO	96.10%	95.58%	97.59%	97.70%
Proposed Method	98.28%	97.90%	98.49%	98.97%

TABLE VI. CLASSIFICATION EFFICIENCY COMPARISON FOR NIM CLASSIFICATION WITH DEEP LEARNING MODELS

Method	Accuracy	Precision	Recall	F1-Measure
AlexNet	92.30%	95.33%	92.38%	94.30%
GoogleNet	93.48%	94.91%	93.55%	95.10%
VGGNet	91.66%	93.43%	92.51%	93.91%
ResNet	95.76%	97.80%	95.41%	95.61%
YOLO	96.81%	98.51%	98.69%	91.90%
Proposed Method	98.90%	99.65%	99.41%	99.38%

TABLE VII. CLASSIFICATION EFFICIENCY COMPARISON FOR SCC CLASSIFICATION WITH DEEP LEARNING MODELS

Method	Accuracy	Precision	Recall	F1-Measure
AlexNet	90.41%	90.28%	91.70%	92.90%
GoogleNet	93.81%	91.83%	90.33%	91.80%
VGGNet	91.94%	91.48%	92.63%	93.18%
ResNet	90.18%	91.90%	92.40%	92.70%
YOLO	92.30%	93.63%	94.64%	94.65%
Proposed Method	97.15%	96.79%	97.24%	96.90%

Table IV showcases the performance comparison of the proposed method with popular deep learning models for HSIL classification using multicell pap smear images. The accuracy of the proposed method is demonstrated to be 98.15%, which outperforms other models such as AlexNet (91.25%), GoogleNet (93.28%), VGGNet (94.00%), ResNet (94.60%), and YOLO (96.72%). Table V presents the comparison results for LSIL classification using multicell pap smear images. The proposed method again demonstrates superior performance, achieving an accuracy of 98.28%. This accuracy surpasses the other models, including AlexNet (91.72%), GoogleNet (92.71%), VGGNet (93.41%), ResNet (94.99%), and YOLO (96.10%). Table VI illustrates the comparison of NIM classification efficiency using multicell pap smear images. Once more, the proposed method stands out with an accuracy of 98.90%, outperforming AlexNet (92.30%), GoogleNet (93.48%), VGGNet (91.66%), ResNet (95.76%), and YOLO (96.81%). The comparison for SCC classification using multicell Pap smear images is presented in Table VII. The proposed method exhibits an accuracy of 97.15%, surpassing AlexNet (90.41%), GoogleNet (93.81%), VGGNet (91.94%), ResNet (90.18%), and YOLO (92.30%). Similar to previous comparisons, the proposed method achieves higher values in Precision, Recall, and F1-Measure, underscoring its effectiveness in classifying SCC using multicell pap smear images. The analysis of the comparison tables demonstrates that the proposed methodology consistently outperforms existing deep learning models across all stages of cervical cancer classification, including HSIL, LSIL, NIM, and SCC.

C. Performance Comparison with Existing Methods

The proposed methodology addresses a critical challenge in cervical cancer classification by focusing specifically on multicell Pap smear images, where existing methods tend to show decreased accuracy. This section presents a comprehensive performance comparison between the proposed method and several existing techniques.

TABLE VIII. CLASSIFICATION EFFICIENCY COMPARISON FOR HSIL CLASSIFICATION WITH EXISTING METHODS

Method	Accuracy	Precision	Recall	F1-Measure
Md Mamunur Rahaman et al. [14]	86.25%	90.69%	86.9%	88.89%
Alyafeai et al. [16]	88.28%	91.70%	87.60%	89.60%
Yuexiang Li et al. [17]	89.00%	92.20%	90.10%	91.10%
Deepa. K et al. [20]	89.60%	93.70%	92.60%	92.70%
Proposed Method	98.15%	98.52%	97.33%	98.43%

TABLE IX. CLASSIFICATION EFFICIENCY COMPARISON FOR LSIL CLASSIFICATION WITH EXISTING METHODS

Method	Accuracy	Precision	Recall	F1-Measure
Md Mamunur Rahaman et al. [14]	87.13%	90.73%	91.36%	90.55%
Alyafeai et al. [16]	88.07%	91.77%	90.86%	90.81%
Yuexiang Li et al. [17]	88.73%	92.60%	92.90%	92.75%
Deepa. K et al. [20]	90.24%	93.70%	92.07%	92.94%
Proposed Method	98.28%	97.90%	98.49%	98.97%

TABLE X. CLASSIFICATION EFFICIENCY COMPARISON FOR NIM CLASSIFICATION WITH EXISTING METHODS

Method	Accuracy	Precision	Recall	F1-Measure
Md Mamunur Rahaman et al. [14]	87.30%	90.33%	87.38%	90.30%
Alyafeai et al. [16]	88.98%	91.91%	88.95%	91.10%
Yuexiang Li et al. [17]	86.16%	88.43%	87.51%	88.91%
Deepa. K et al. [20]	90.76%	92.80%	90.41%	91.61%
Proposed Method	98.90%	99.65%	99.41%	99.38%

TABLE XI. CLASSIFICATION EFFICIENCY COMPARISON FOR SCC CLASSIFICATION WITH EXISTING METHODS

Method	Accuracy	Precision	Recall	F1-Measure
Md Mamunur Rahaman et al. [14]	84.29%	84.10%	85.38%	84.78%
Alyafeai et al. [16]	88.23%	87.34%	87.02%	87.18%
Yuexiang Li et al. [17]	85.58%	85.74%	86.15%	85.94%
Deepa. K et al. [20]	84.46%	85.68%	85.88%	85.78%
Proposed Method	97.15%	96.79%	97.24%	96.90%

Table VIII highlights the comparison of the proposed method with existing techniques for HSIL classification. Notably, the proposed method achieves an accuracy of 98.15%, outperforming other methods such as Md Mamunur Rahaman et al. (86.25%), Alyafeai et al. (88.28%), Yuexiang Li et al. (89.00%), and Deepa K et al. (89.60%). In Table IX, the performance comparison for LSIL classification is presented. The proposed method once again achieving an accuracy of 98.28%. This accuracy significantly surpasses existing methods such as Md Mamunur Rahaman et al. (87.13%), Alyafeai et al. (88.07%), Yuexiang Li et al. (88.73%), and Deepa. K et al. (90.24%). Table X showcases the comparison results for NIM classification. The proposed method excels with an accuracy of 98.90%, surpassing Md Mamunur Rahaman et al. (87.30%), Alyafeai et al. (88.98%), Yuexiang Li et al. (86.16%), and Deepa K et al. (90.76%). The performance comparison for SCC classification is provided in Table XI. The proposed method achieves an accuracy of 97.15%, showcasing its superiority over Md Mamunur Rahaman et al. (84.29%), Alyafeai et al. (88.23%), Yuexiang Li et al. (85.58%), and Deepa K et al. (84.46%). The performance comparison clearly illustrates the proposed method's exceptional capability in addressing the limitations of existing methods, which struggle with decreased accuracy in multicell pap smear images.

The proposed method achieves remarkable accuracy, precision, recall, and F1-Measure values, indicating its robustness in accurately identifying different cell types and cancer stages using multicell Pap smear images. The existing methods were limited by their use of single pap smear images. This can be a problem because single pap smear images may not contain enough information to accurately classify the cell type or cancer stage. According to the results, the proposed method addresses this limitation by using multicell Pap smear images.

D. Training Efficiency Analysis

The training phase of deep learning models holds paramount importance in achieving accurate and robust results. To ensure optimal training, a combination of advanced training algorithms and techniques was employed. In this research, the Adam optimizer was selected as the primary algorithm for training the model. The Adam optimizer is known for its efficiency in parameter optimization and model tuning, making it well-suited for the complex and high-dimensional nature of deep learning models. To comprehensively evaluate the performance of the proposed model, a five-fold cross-validation strategy was adopted. This approach enhances the model's generalization ability by partitioning the dataset into five subsets, with four subsets utilized for training and one subset reserved for testing. The use of cross-validation mitigates the risk of overfitting and provides a robust estimate of the model's performance on unseen data. The average testing accuracy emerged as the pivotal metric to gauge the effectiveness of the proposed cervical cancer classification system. By aggregating the accuracy results from each fold, a more comprehensive and reliable measure of the model's classification performance was obtained. The training efficiency and effectiveness of the proposed method are visually depicted in Fig. 6 to Fig. 10 These figures show the model's performance across different folds, providing insights into its consistency and stability.

E. Computational Efficiency Analysis

This section has two sections: the first section analyses the training time analysis, and the second section analyses the testing time analysis.

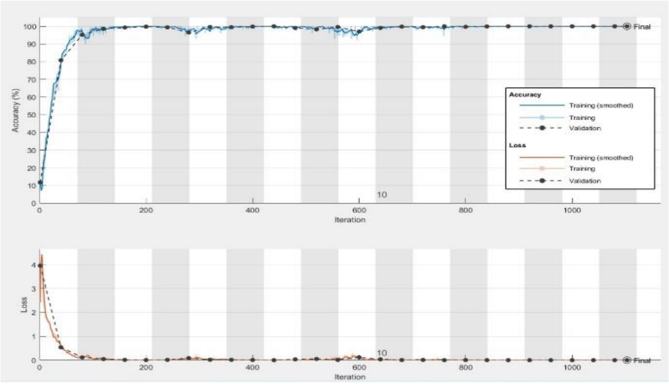


Fig. 6. Training details of cross fold 1.

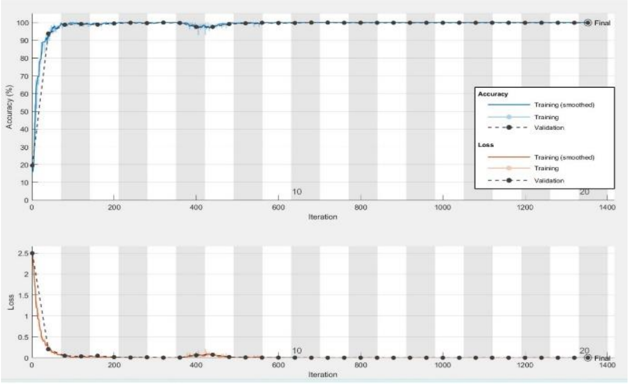


Fig. 7. Training details of cross fold 2.

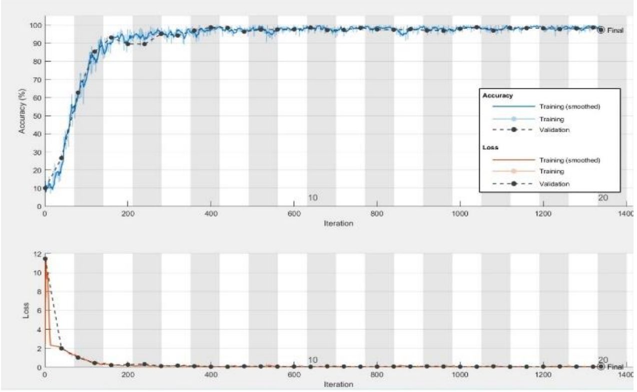


Fig. 8. Training details of cross fold 3.

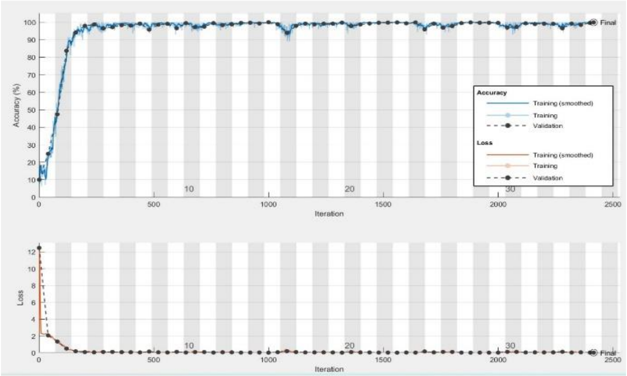


Fig. 9. Training details of cross fold 4.

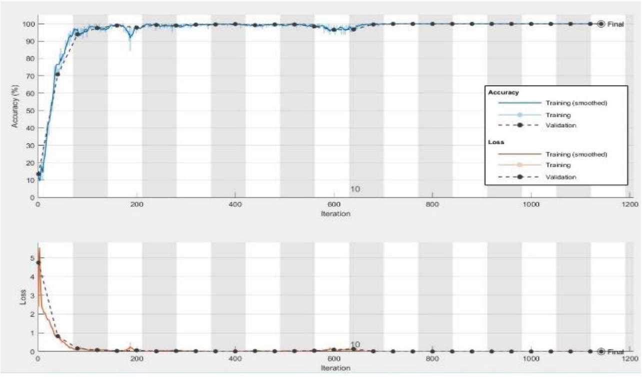


Fig. 10. Training details of cross fold 5.

The evaluation of computational efficiency plays a pivotal role in assessing the practical feasibility of deep learning models for medical image analysis. The efficiency of such models is influenced by a multitude of factors, including the effectiveness of the segmentation algorithm, the feature extraction model, the effectiveness of the training algorithm, the image type, and the image size. In order to conduct a thorough comparison of training time efficiency, three distinct sets of Pap smear images were utilized: the first set consisting of 200 images, the second containing 400 images, and the third encompassing 600 images. Each set was utilized for training the proposed deep learning model, thereby allowing for a comprehensive assessment of training time variations across different dataset sizes. The training process was carried out using the computational resources detailed in Section IV (B).

Table XII explains the training efficiency comparison results. This proposed method is much faster than existing methods. For example, it takes only 13.65 minutes to train on Dataset 1, while AlexNet, GoogleNet, VGGNet, and ResNet take 105-147 minutes. This trend continues for Dataset 2 and Dataset 3. Existing methods are effective, but they take too long to train, which is not practical for clinical applications. This proposed method is much faster without sacrificing accuracy. It uses Graph-Cut Guided ROI Segmentation and the enhanced DenseNet architecture to achieve this.

In addition to training time, the classification time of a cervical cancer classification methodology is a crucial performance metric, particularly for real-time clinical applications. This section presents an analysis of the classification time required by various methodologies, including the proposed approach, when processing images of different sizes. The assessment is carried out using three images of varying sizes: Image 1 (800 kb), Image 2 (1200 kb), and Image 3 (1800 kb). The following table presents the classification times for each methodology when applied to the three different images.

TABLE XII. TRAINING EFFICIENCY COMPARISON RESULTS

Methods	Dataset 1 (200 images)	Dataset 2 (400 images)	Dataset 3 (600 images)
AlexNet	147 minutes	241.5 minutes	329.25 minutes
GoogleNet	131.7 minutes	227.05 minutes	321.8 minutes
VGGNet	126 minutes	214.6 minutes	288 minutes
ResNet	105 minutes	171.7 minutes	213.8 minutes
YOLO	84 minutes	154.4 minutes	201.0 minutes
Md Mamunur Rahaman et al.	60.9 minutes	93.1 minutes	148.7 minutes
Alyafeai et al.	52.5 minutes	81.1 minutes	129.25 minutes
Yuexiang Li et al.	63 minutes	100.8 minutes	135.75 minutes
Deepa. K et al.	71.4 minutes	95.25 minutes	142.25 minutes
Proposed Method	13.65 minutes	17.85 minutes	32.55 minutes

This method is fast at classifying images of all sizes. For example, it takes 0.32 seconds to classify an image of 800 kb, 0.6475 seconds to classify an image of 1200 kb, and 1.04 seconds to classify an image of 1800 kb. This is important because it means that the proposed method can be used in real-time clinical applications. Other methods, such as GoogleNet, ResNet, and YOLO, are slower at classifying images of larger sizes. Alyafeai et al.'s method is fast for small images, but it becomes slower for larger images. The efficiency of proposed method is due to the Graph-Cut Guided ROI Segmentation Algorithm and the enhanced lightweight DenseNet architecture. These two techniques help to reduce the training and classification times while still ensuring accurate cervical cancer classification.

TABLE XIII. TESTING EFFICIENCY COMPARISON RESULTS

Methods	Image 1 (800 kb)	Image 2 (1200 kb)	Image 3 (1800 kb)
AlexNet	0.375 seconds	0.0975 seconds	0.4125 seconds
GoogleNet	0.90 seconds	1.80 seconds	2.70 seconds
VGGNet	0.36 seconds	0.5475 seconds	0.685 seconds
ResNet	0.80 seconds	1.60 seconds	2.40 seconds
YOLO	0.72 seconds	1.32 seconds	1.92 seconds
Md Mamunur Rahaman et al.	0.0925 seconds	0.6375 seconds	0.975 seconds
Alyafeai et al.	0.1425 seconds	0.3475 seconds	0.6825 seconds
Yuexiang Li et al.	0.61 seconds	1.07 seconds	1.53 seconds
Deepa. K et al.	0.67 seconds	1.12 seconds	1.68 seconds
Proposed Method	0.32 seconds	0.6475 seconds	1.04 seconds

F. Discussions

Cervical cancer is a major global health concern, and the accurate classification of cervical cells plays a critical role in its early detection and effective treatment. Many existing approaches to cervical cancer classification use separate models for segmentation and classification. This approach not only escalates computational demands but also prolongs training and inference times. The proposed method addresses this issue by integrating segmentation and classification into a single pipeline. The Graph-Cut Guided ROI Segmentation Algorithm and the enhanced lightweight DenseNet architecture enable simultaneous segmentation and feature extraction, reducing computational complexity and speeding up the overall process.

Existing approaches often rely on feature fusion to combine the outputs of separate models. However, this approach introduces complexities and potential sources of error. Proposed method avoids this pitfall by eliminating the need for feature fusion altogether. Instead, proposed unified architecture inherently incorporates both segmentation and feature extraction, mitigating the need for complicated fusion strategies. This simplification reduces the risk of errors introduced through fusion methods, leading to a more robust and accurate cervical cancer classification system.

Most existing methods for cervical cancer classification are tested with single cell images, which does not reflect the reality of pap smear images that often contain multiple cells. The proposed method directly addresses this limitation by utilizing multicell Pap smear images. The Graph-Cut Guided ROI Segmentation Algorithm effectively handles multiple cells within an image, overcoming the challenges of accurate segmentation and classification in such scenarios. This adaptability enhances the method's applicability to real-world situations, improving its reliability and clinical utility.

The results of this study demonstrate the exceptional performance of the proposed method in cervical cancer classification using multicell pap smear images. Proposed method consistently outperforms existing deep learning models and methodologies across all stages of cervical cancer classification. The accuracy, precision, recall, and F1-measure values attained underscore the method's efficacy in accurately identifying different cell types and cancer stages.

In addition, the proposed method is more efficient for training because it does not need to train multiple models or use complex feature fusion techniques. This means that it takes less time and uses fewer resources to train the model, while still being able to classify images accurately. Proposed method has been shown to be more effective than existing deep learning models for classifying cervical cancer at all stages. It is also faster and more accurate at classifying images of varying sizes. This is because proposed method uses a unified architecture that eliminates the need for complex fusion strategies.

While proposed method shows promising results, there are certain limitations that need to be acknowledged. Firstly, the performance of the method heavily relies on the quality of the input images. Variability in image quality, lighting conditions, and staining techniques can affect the accuracy of segmentation and classification. Secondly, although proposed method improves computational efficiency compared to existing methods, there is still room for optimization to further enhance speed and resource utilization. Lastly, the evaluation of the proposed method is based on specific datasets, and its generalization to diverse populations and different acquisition conditions should be thoroughly validated before widespread clinical adoption.

V. Conclusion

This study presents a novel and effective approach for automated cervical cancer classification using multicell pap smear images. The method is designed to address the limitations of existing approaches and enhance the accuracy, efficiency, and practicality of cervical cancer diagnosis. The implications of these findings are profound for the medical community. This research not only highlights the potential of deep learning in transforming cervical cancer diagnosis but also addresses the critical need for efficiency in real-world applications. The successful fusion of the Graph-Cut Guided ROI Segmentation Algorithm and the enhanced lightweight DenseNet architecture demonstrates the possibility of accurate and rapid classification, even in the context of multicell pap smear images. This has the potential to enhance clinical decision-making, treatment planning, and patient outcomes.

References

[1] S. Pimple and G. Mishra, ‘Cancer cervix: Epidemiology and disease burden’, Cytojournal, vol. 19, no. 21, pp. 21, Mar. 2022.

[2] Wang, C.-W., Liou, Y.-A., Lin, Y.-J., Chang, C.-C., Chu, P.-H., Lee, Y.-C., Wang, C.-H., & Chao, T.-K. (2021), “Artificial intelligence-assisted fast screening cervical high grade squamous intraepithelial lesion and squamous cell carcinoma diagnosis and treatment planning,” Scientific Reports, vol. 11, no. 1, Aug. 2021.

[3] K. P. Win, Y. Kitjaidure, K. Hamamoto, and T. M. Aung, “Computer-Assisted screening for cervical cancer using digital image processing of pap smear images,” Applied Sciences, vol. 10, no. 5, p. 1800, Mar. 2020.

[4] A. Gupta, A. Parveen, A. Kumar, and P. Yadav, ‘Advancement in deep learning methods for diagnosis and prognosis of cervical cancer’, Curr. Genomics, vol. 23, no. 4, pp. 234–245, Aug. 2022.

[5] A. Anaya-Isaza, L. Mera-Jiménez, and M. Zequera-Diaz, “An overview of deep learning in medical imaging,” Informatics in Medicine Unlocked, vol. 26, pp. 100723, Jan. 2021.

[6] M. Illimoottil and D. T. Ginat, “Recent advances in deep learning and medical imaging for head and neck cancer treatment: MRI, CT, and PET scans,” Cancers, vol. 15, no. 13, pp. 3267, Jun. 2023.

[7] N. Youneszade, M. Marjani and C. P. Pei, "Deep Learning in Cervical Cancer Diagnosis: Architecture, Opportunities, and Open Research Challenges," in IEEE Access, vol. 11, pp. 6133-6149, 2023.

[8] S. Dash, P. K. Sethy, and S. K. Behera, “Cervical transformation zone segmentation and classification based on improved inception-ResNet-V2 using colposcopy images”, Cancer Inform., vol. 22, pp. 1-8, Mar. 2023.

[9] N. Nanthini, B. Kaviya Sree, M. Kavya and K. Monika, "Cervical Cancer Cell Segmentation and Classification using ML Approach," In: Proc. of 7th International Conf. Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp. 1090-1095.

[10] H. Alquran, M. Alsalatie, W. A. Mustafa, R. A. Abdi, and A. R. Ismail, “Cervical Net: A novel cervical cancer classification using feature fusion,” Bioengineering, vol. 9, no. 10, pp. 578, 2022.

[11] S. Shinde, M. Kalbhor, and P. Wajire, “DeepCyto: a hybrid framework for cervical cancer classification by using deep feature fusion of

cytology images", Math. Biosci. Eng., vol. 19, no. 7, pp. 6415–6434, Apr. 2022.

[12] A. Ghoneim, G. Muhammad, and M. S. Hossain, "Cervical cancer classification using convolutional neural networks and extreme learning machines," Future Generation Computer Systems, vol. 102, pp. 643–649, Jan. 2020.

[13] N. Youneszade, M. Marjani and C. P. Pei, "Deep Learning in Cervical Cancer Diagnosis: Architecture, Opportunities, and Open Research Challenges," in IEEE Access, vol. 11, pp. 6133-6149, 2023.

[14] Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Wu, X., Li, X., & Wang, Q , "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," Computers in Biology and Medicine, vol. 136, pp. 104649, Sep. 2021.

[15] P. Sukumar and S. Ravi, "Computer aided detection and classification of Pap smear cell images using principal component analysis", Int. J. Bio-inspired Comput., vol. 11, pp. 257, 2018.

[16] Z. Alyafeai and L. Ghouti, "A fully-automated deep learning pipeline for cervical cancer classification," Expert Systems with Applications, vol. 141, pp. 112951, 2020.

[17] C. Tang, J. Chang, C. Chu, K. Ma, Q. Li, Y. Zheng, "Computer-aided cervical cancer diagnosis using time-lapsed colposcopic images", IEEE Trans. Med. Imaging, vol. 39, no. 11, pp. 3403–3415, Nov. 2020.

[18] L. D. Nguyen, R. Gao, and D. Lin, "Biomedical image classification based on a feature concatenation and ensemble of deep CNNs", J Ambient Intell Human Comput, vol. 9, no. 10, pp. 578-591, 2019.

[19] S. Gautam, H. K. K., N. Jith, A. K. Sao, A. Bhavsar, and A. Natarajan, "Considerations for a PAP smear image analysis system with CNN features", arXiv [cs.CV], 23-Jun-2018.

[20] K. Deepa, "PAP smear Image Classification to Predict Urinary Cancer Using Artificial Neural Networks", Annals of the Romanian Society for Cell Biology, vol. 25, no. 2, pp. 1092–1098, 2021.

[21] C. Li, H. Chen, L. Zhang, N. Xu, D. Xue, Z. Hu, H. Ma, and H. Sun, "Cervical Histopathology Image Classification Using Multilayer Hidden Conditional Random Fields and Weakly Supervised Learning," in IEEE Access, vol. 7, pp. 90378-90397, 2019.

[22] E. Hussain, 'Liquid based cytology pap smear images for multi-class diagnosis of cervical cancer'. Mendeley, 2019.

[23] A. Alrajjal, V. Pansare, M. S. R. Choudhury, M. Y. A. Khan, and V. B. Shidham, 'Squamous intraepithelial lesions (SIL: LSIL, HSIL, ASCUS, ASC-H, LSIL-H) of Uterine Cervix and Bethesda System', Cytojournal, vol. 18, no. 16, pp. 16, Jul. 2021.

[24] G. F. C. Campos, S. M. Mastelini, G. J. Aguiar, R. G. Mantovani, L. F. De Melo, and S. Barbon, "Machine learning hyperparameter selection for Contrast Limited Adaptive Histogram Equalization," Eurasip Journal on Image and Video Processing, vol. 2019, no. 1, May 2019.

[25] Shorten, C., Khoshgoftaar, T.M, "A survey on Image Data Augmentation for Deep Learning", J Big Data, vol. 6, pp. 60, 2019.

[26] Tao Zhou, XinYu Ye, HuiLing Lu, Xiaomin Zheng, Shi Qiu, and YunCan Liu, "Dense Convolutional Network and Its Application in Medical Image Analysis", BioMed Research International, vol. 2022, pp 22, 2022.

[27] N. Hasan, Y. Bao, A. Shawon, and Y. Huang, "DenseNet Convolutional Neural Networks Application for predicting COVID-19 using CT Image," SN Computer Science, vol. 2, no. 5, Jul. 2021.

Advanced Techniques for Recognizing Emotions: A Unified Approach using Facial Patterns, Speech Attributes, and Multimedia Descriptors

Kummari Ramyasree[1], Chennupati Sumanth Kumar[2]
Department of E & ECE, GITAM Deemed to be University, Visakhapatnam -530045, AP, India[1, 2]
Department of ECE, Guru Nanak Institutions Technical Campus, Hyderabad -501506, Telangana, India[1]

***Abstract*—The inability to efficiently store distinguishing edges, local appearance-based textured descriptions generally have limited performance in detecting facial expression analysis. The existing technology has certain drawbacks, such as the potential for poor edge-related disturbance in face photos and the reliance on present sets of characteristics that might fail to adequately represent the subtleties of emotions and thoughts in a variety of contexts. In order to overcome the difficulties associated with identifying facial expressions identification and emotion categorization, this study presents an innovative structure that combines three different information sets: a new multimedia descriptors, prosodic functions, and Local Differential Pattern (LDP). The principal driving force is the existence of noise-induced warped and weak edges in face pictures, which result in inaccurate expressions characteristic assessment. By identifying and encoding only greater edge reactions, as opposed to standard local descriptors that the LDP approach improves the endurance of face feature extraction. Robinson Compass and Kirsch Compass Masks are used for recognising edges, and the LDP formulation encodes each pixel with seven bits of information to reduce code repetition. The last category comprises Long-Term Average Spectrum (LTAS) obtained from signals related to speech, Mel-Frequency Cepstral Coefficients (MFCC), and Forming agents. Fisher Criterion is used to reduce dimensionality, and unpredictable characteristics are used in picking features. Emotion prediction is achieved by classifying two distinct circumstances using Support Vector Machine (SVM) and Decision Tree (DT) algorithms, and combining the obtained data. The research also presents a unique Visual or audio Descriptors that gives priority to key structure selections and face positioning for Audio-visual input. A concise depiction of expression is offered by the suggested Self-Similarity Distance Matrix (SSDM), which uses facial highlight points to estimate both time and space correlations. Formant frequency range, energy sources, probabilistic properties, and spectroscopic aspects define the acoustic signal. The 98% accuracy rate is attained by the emotion recognition algorithm. Major improvements upon cutting-edge techniques are shown in validation studies on the SAVEE and RML information sets, highlighting the usefulness of the suggested model in identifying and categorising emotions and facial movements in a variety of contexts. The implementation of this research is done by using Python tool.**

Keywords*—Local Difference Pattern (LDP); Mel-Frequency Cepstral Coefficients (MFCC); Long-Term Average Spectrum (LTAS); Self-Similarity Distance Matrix (SSDM); Support Vector Machine (SVM)

I. Introduction

Emotion recognition is an intricate yet essential component of human communication, and computing and AI has placed a great deal of emphasis on it recently. The capacity to precisely recognise and comprehend human feelings has broad applications, from boosting behavioural care providers to advancing consumer-computer connection [1]. Facial expressions are an inherent means for people to communicate a wide range of emotions, and algorithmic systems may be developed to read similar nuanced signs [2]. Basic features including mouth expressions, blinking eyes and eyebrows twitches may be detected by facial identification applications, which can then be used to deduce moods like surprise, pleasure, sorrow, and rage. Apart from expressions on the face, speech assessment is an essential component in the identification of emotions [3][4]. Another approach used in mood detection systems is identification of gestures. Human behaviour, which includes posture and movements, may give important information about how one feels. Huge collections of movement of the body training may be used to train machine learning algorithms to recognise certain gestures as indicative of certain feelings [5]. This method is especially useful in situations when it may be difficult to read or understand facial emotions, including in crowded spaces or through video conferences. The creation of emotions detection mechanisms has great potential for a variety of uses as technology evolves. The capacity of computers to comprehend and react to human emotions offers up fresh prospects for developing empathic and flexible innovations, from interactions between humans and computers to mental wellness assessment and personalised experiences for consumers [6]. Still, the significance of appropriate research and implementation in such developing sector is underscored by ethical issues, privacy problems, and the possibility of bias in identifying feelings systems.

Numerous methodologies are now in use for recognising feelings, and they all use different forms and ways to gather and process emotional information. Among the techniques that have been studied and used frequently is facial expressions assessment. This method entails the detection and analysis of facial reactions and motions employing machine learning tools. Convolutional Neural Networks (CNNs) are a popular neural network method used by investigators and programmers to recognise similarities linked to different feelings from enormous datasets of labelled facial movements

[7]. By examining the meaning included in words written or spoken, machine learning techniques improve this technique even further and make it possible to identify the feelings communicated via language. Voice-enabled devices, digital assistants, and contact centres are among the places where identifying emotions in conversation is frequently used [8]. The field of gesture identification is centred on the interpretation of nonverbal cues such as posture, movements of the hands, and various other physiological motions. This technique is especially useful in situations when facial emotions might not be apparent or may lack sufficient data. Machine learning techniques may be developed on databases that link particular movements to associated sentiments [9]. Such algorithms frequently depend on neural networks with recurrence or related sequence-based designs. Applications such as augmented realities, surveillance footage, and interactions between humans and computers can benefit from this strategy. The use of multiple methods techniques work especially well in everyday life when there are a variety of influenced by context signs of emotion [10]. Even though these techniques have shown promise in a number of applications, problems still exist. Important factors to take into account are the possibility of biased results, confidentiality difficulties, and ethical issues.

A significant barrier to attaining global application is the subtleties and cultural differences in expressing one's feelings [11]. Dependent on environment is a further important restriction. Because feelings are so dependent on the setting, current frameworks may find it difficult to account for the subtle differences in how emotions seem in various contexts. The subjective nature and variability among individuals present further difficulties for emotion detection methods trying to be generalised [12]. Individualised variations in personality, socioeconomic situation, or mental state can provide unpredictability that is challenging for systems to compensate for, and humans may exhibit moods in various manners. The use of such devices in vulnerable or public areas may violate people's fundamental right to safety, hence rules and moral requirements must be carefully considered in order guarantee proper usage [13].

Recent years have seen tremendous progress in the field of emotion recognition, especially with the incorporation of multimedia descriptors, speech characteristics, and facial patterns. Despite these advancements, there are still significant obstacles in the way of the practical uses of current techniques. A significant problem is the restricted capacity to represent the subtleties of affective responses in many media. The process of feature extraction, which is essential for distinguishing minute differences in speech intonations, facial expressions, and multimedia information, is frequently difficult for traditional approaches. The suggested strategy addresses the drawbacks of current techniques by introducing a number of novel algorithms in response to these difficulties. We highlight how important it is to extract features using the Local Directional Pattern (LDP) method, which improves facial pattern representation and guarantees a deeper examination of emotional indicators. But the problem goes beyond feature extraction, which is a precisely sophisticated method like feature selection and dimensionality reduction are needed. In order to ensure a more effective and efficient emotion detection procedure, these stages are essential for simplifying the data and conserving just the most discriminative aspects.

The limitations of traditional classification techniques are addressed by our approach's use of Decision Tree Classifier and Support Vector Machines (SVM). These classifiers improve the system's capacity to identify intricate patterns in multimedia information, speech tones, and facial expressions, enabling a more complicated comprehension of emotions. Moreover, the incorporation of the Self-Similarity Distance Matrix (SSDM) enhances the analysis and facilitates a thorough assessment of similarity patterns among modalities. A fundamental problem with current emotion identification systems is the absence of a cohesive methodology that effectively integrates data from many sources. Seeing this gap, the foundation of our strategy is our suggested Emotion Recognition Fusion mechanism. This fusion technique offers a comprehensive and more accurate representation of the subject's emotional state by cleverly merging information from multimedia descriptors, voice features, and facial patterns. The key contributions of the proposed framework for Facial Expression Recognition and emotion classification can be summarized as follows:

1) The system offers a complete method to capture subtle emotions by integrating three different feature sets: prosodic characteristics, a unique Audio-Visual Descriptor, and Weighted Edge Local Directional Pattern (WELDP). This lessens the reliance on present feature sets.

2) WELDP optimises the encoding process by using seven bits to represent each pixel, hence minimising code repetition.

3) This helps to portray face characteristics more effectively and efficiently, especially when there are weak or warped edges present.

4) The suggested system leverages both individual models by combining the Support Vector Machine and Decision Tree methods for categorization.

5) Combining the output from various models improves the prediction of emotions and yields a categorization result that is more trustworthy.

6) A novel method is provided by the introduction of the Audio-Visual Descriptor based on the Self-Similarity Distance Matrix (SSDM).

7) SSDM computes spatial and temporal distances using facial landmark points, giving priority to face alignment and key frame selection for Audio-Visual input.

8) This results in a succinct yet powerful depiction of emotion in a variety of settings.

The format for the enduring paragraphs is as follows: The relevant work based on various methodologies for diabetes prediction is examined in Section II, and the research gap is identified in Section III. The feature selection and classification process for the proposed method is explained in the Section IV. The outcomes and considerations are covered in Section V; the prospective applications for the future are covered in Section VI.

II. Related Works

The techniques for identifying emotions using multiple interfaces EEG data and multipurpose physiological indicators are the main topic of this extensive literature analysis [14]. The research employs a conventional emotional identification pipeline, looking at different approaches to extracting features like wavelet transformed and nonlinear behaviour, and also decreasing features and machine learning (ML) classification system design methodologies like k-nearest neighbour (KNN), naive Bayesian (NB), support vector machine (SVM), and random forest (RF). The work delves deeper into the complex relationship between various brain regions and mental conditions by analysing EEG patterns that are significantly connected with sentiments. The paper also compares and contrasts machine learning and deep learning methods for emotion identification, highlighting the advantages and disadvantages of each. The study finds that the scope of the collection of features and the data set used for training have a major impact on how well DL models can recognise emotions.

Communication between humans and computers has greatly increased availability of educational resources, data, and the sharing of significant skills on a worldwide scale. The model in [15] offered suggests fusing speech qualities and face emotions to overcome these drawbacks. The technique particularly employs the Speech (Mel Frequency Cepstral Coefficients) and Facial (Maximally Stable Extremal Regions) aspects, which add to an organised and methodical investigation. This methodology presents a more resilient and adaptable emotion identification system, highlighting the possibility of using multifaceted characteristics for enhanced accuracy in a variety of situations in life. Although the bi-modal approach which combines voice and face features has shown to improve accuracy and resilience in the sense of emotion identification network when compared with unimodal. This method, one significant drawback is the requirement for additional expansion.

Two neural network methods in [16] for recognising and categorising objects are included in the suggested flexible design. These approaches are taught separately to facilitate use in real time. AdaBoost cascading models are used for face identification, and then neighbourhood difference features (NDF) are extracted to get localised attractiveness knowledge. This allows for the development of various structures depending on the interactions between surrounding areas. The adaptable structure of the system, which focuses on seven primary facial gestures, enables it to be extended to categorise a wide variety of emotions. In order to manage mis-/false detection, a classifier using random forests with a deep mood indicator has been developed, providing self-determination through sexual orientation and face complexion. The investigation of shape characteristics and face modification brings complexities that can necessitate advanced mathematical methods, which could have an impact on the computing effectiveness of immediate video analysis.

This study in [17] offers a thorough overview of micro-expression evaluation using videos, highlighting the distinctions among broad and tiny movements and using them as a foundation for assessment. The research presents a fresh collection of data, the micro-and-macro emotion warehouse (MMEW), with a greater quantity of video frames and tagged groups of emotions in recognition of the shortcomings of the current micro-expression databases. The writers compare typical techniques uniformly on CAS(ME)2 for finding and on MMEW and SAMM for classification. In the latter section of this investigation, several avenues for further study are discussed, emphasising the dynamic environment and continuous difficulties in the field of video-based micro-expression evaluation. Quick breakthroughs or changes might arise from the multifaceted nature of micro-expression research and the ever-changing nature of technological devices, requiring regular updates exceeding the scope of this paper. To get latest findings in this quickly increasing field of research, investigators should be aware of how the environment is changing and take into account further sources.

By creating an in-depth structure [18] that combines three different classifiers a deep neural network (DNN), a convolutional neural network (CNN), and a recurrent neural network (RNN), this study tackles the difficult job of audio emotion detection. The approach uses segment-level mel-spectrograms (MS), frame-level modest descriptions (LLDs), and utterance-level outcomes of the highest-level statistical algorithms (HSFs) on LLDs to concentrate on categorising detection of four different feelings. Adopting a multifaceted learning approach, the separate versions of LLD-RNN, MS-CNN, and HSF-DNN that are produced are merged to conduct extraction on continual emotion characteristics and categorise defined types of emotions all at once. It is important to take into account the computing requirements of these optimisation techniques, especially if working with huge data sets or apps that operate in real time. This highlights the necessity for efficient and flexible systems in further research.

The promise of blended learning for automated identification of emotions is discussed in this work, with a focus on the relationships and dependencies between sight and aural domains that are still poorly understood. This approach [19] to Multimodal Emotion Recognition Metric Learning (MERML) seeks to simultaneously learn an adequate representational in a space known as latent and a selective score for both techniques, appreciating the distinctive features of each. A Support Vector Machine (SVM) kernel founded on the Radial Basis Function (RBF) effectively applies the learnt metric. The work acknowledges that developing an efficient measure in multi-modal settings is an important aim for a variety of use cases involving machine learning. There are worries over adaptability across different datasets with different channels and emotions as the evaluation's statistics' unique qualities and complexity may have an impact on MERML's efficacy. For assessing MERML's realistic practicality in cases outside of the assessed datasets, more research is necessary to test its effectiveness and versatility in applications that utilise real-time or huge data sets. Even if this research shows better performance, there is still work to be done on the accessibility of the learnt measure and its applicability to other multimodal recognition of emotions activities.

The literature addresses traditional emotional identification pipelines, like subject-independent behavioural characteristics and machine learning classification algorithms, but also acknowledges the field's developing problems. A possible solution is the suggested bi-modal strategy that combines speech and facial clues; nonetheless, the study rightly highlights the need for further research, especially when dealing with a variety of human faces and environments. Moreover, the study of face expression recognition in autonomous vehicles and human-computer interactions highlights the necessity for thorough assessment, particularly with the influence of temporal and geographical factors. The study on micro-expression evaluation emphasizes the dynamic nature of technology in the sector and emphasizes the significance of dataset limits. While highlighting the need for more research and highlighting the difficulties of complicated models, the work on audio emotion recognition presents intriguing multimodal learning approaches. In a similar vein, while Multimodal Emotion Recognition Metric Learning (MERML) research shows good results, more investigation is necessary to ensure that the model is flexible enough to work with different datasets and in real-world circumstances. In summary, these assessments of the literature successfully highlight the deficiencies and intricacies in the field of emotion recognition, offering significant perspectives for further studies.

III. Problem Statement

This comprehensive overview of the literature aims to address the central problem of identifying mental states using different interfaces, with particular focus on EEG results and numerous physiological markers. The research explores conventional emotional identification pipelines, which include smaller feature sizes and classification algorithms such as KNN, NB, SVM, and RF, in addition to methods like wavelet modification and nonlinear behaviour extraction. Through examining EEG patterns linked to affect, the study investigates the complex relationship between mental states and particular brain regions [14]. It also clarifies the benefits and limitations of using machine learning for emotion recognition as opposed to deep learning methods. The study also addresses the difficulties that arise when applying deep learning techniques in practical settings and emphasizes how important feature selection and data accessibility are in deciding how effective deep learning models are. In order to improve accuracy and overcome the shortcomings of current emotion identification methods, the study suggests a bi-modal approach that combines facial and voice signals [15]. It also explores the rapidly developing field of computational learning for emotion recognition and uses neural network techniques for real-time applications, namely facial expression recognition in HCI, autonomous driving, and micro-expression analysis in movies. Moreover, the paper addresses the complexities of auditory emotion recognition and provides an extensive framework with three different categorization techniques for real-time learning and improved accuracy [18].These difficulties are all related to the larger objective of creating more reliable and effective emotion detection techniques, which emphasizes the critical need for additional study to get beyond present barriers and progress the area.

IV. Proposed SVM and Decision Tree for Emotion Recognition

Datasets from reliable sources, like the Surrey Audio-Visual Expressed Emotion (SAVEE), Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), and Ryerson Multimedia Laboratory (RML), are first acquired for the suggested application. The later steps in the process are built upon these datasets. To ensure consistent feature scaling across various datasets, the data goes through a preprocessing step wherein Min-Max normalization is done. Finding pertinent features in the pre-processed data, the feature extraction procedure applies the Local Directional Pattern (LDP) approach. After that, the Fisher criteria—a technique that finds distinguishing features—are applied to feature selection and dimensionality reduction, maximizing the dataset for further examination. Support vector machines (SVMs) and decision trees are used as classifiers to improve classification accuracy. These classifiers use the chosen features to classify the dataset's emotional content. Furthermore, the Self-Similarity Distance Matrix (SSDM) is utilized as a metric to evaluate the degree of similarity across emotional patterns included in the data. The process of integrating the outputs from various classifiers and matrices is called emotion recognition fusion, and it is the culmination of the suggested technique. By utilizing each component's unique capabilities, this fusion method seeks to increase the overall accuracy and dependability of emotion recognition across the datasets. The all-encompassing strategy, which combines preprocessing, feature extraction, classification, and fusion, highlights how reliable and successful the suggested method, is in identifying emotions in audio-visual data. Fig. 1 explains the conceptual Diagram.

A. Dataset Collection

The data collections utilised in the present investigation are from the Ryerson Multimedia Laboratory (RML), Surrey Audio-Visual Expressed Emotion (SAVEE), and Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). The 1440 audio recordings in the RAVDESS data set are spoken in English by 12 male and 12 female participants. It is composed of eight distinct emotions: fear, fearlessness, rage, calmness, happiness, surprise, disgust, sadness, and neutrality. The 480 videos in the SAVEE data set are narrated in English by four male participants. It is composed of seven distinct emotions: fear, fearless, pleased, astonished, angry, dissatisfied and sad. The 720 videos in the RML data set are voiced in Persian, English, Italian, Urdu, Chinese, and Punjabi [20].

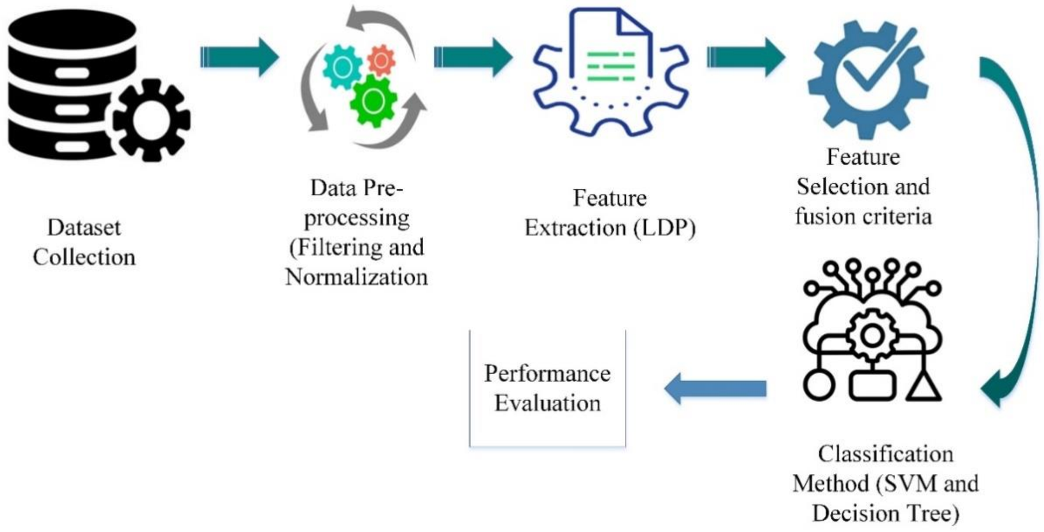

Dataset Collection
Data Pre-processing (Filtering and Normalization
Feature Extraction (LDP)
Feature Selection and fusion criteria
Classification Method (SVM and Decision Tree)
Performance Evaluation

Fig. 1. Overall conceptual diagram.

B. *Data Pre-Processing*

The method of identifying speech and visual disorders is thought to begin with pre-processing. It is used for both pattern extraction and data recognition. Pre-processing techniques are used to modify or amend the speech or visual data, *X(k)*, in order to prepare it for further processing. There is sometimes too much unnecessary and disruptive noise in the transmission of voice or visual data. Traditionally, sound-absorbing cotton or directional microphones were employed to cover up background noise produced by unwanted signals, such as noise from winds and object movement. It is a process for verifying the voice/vision message, *x(k).*If the voice/vision transmission is affected by the surroundings or ambient noise, like *a(k),*this is known as additive disruption. This can be removed using Eq. (1).

$$x(k) = s(k) + a(k) \tag{1}$$

The previous equation yields *s(k),*an exact voice/vision communication. Numerous noise reduction methods can be applied to a noisy speech stream to complete the procedure. The windowing technique: At this stage, the speech or vision information has been split into segments. The windowing functionality, or *w(k),*whose total length is represented by the character l, where l is the audio message's frame length, is used to amplify the signal's frames. Windowing is a type of analysis method where a speech/vision message section in a waveform is multiplied by a time window for a specific form to highlight the signal's intended distinctiveness is shown in Eq. (2).

$$w(k) = 0.54 - 0.46\cos\left(\frac{6.28k}{p-1}\right), \quad 0 \le k \le p-1 \tag{2}$$

Preprocessing must be normalized in order to be categorized. To expedite the learning procedure, the given information has to be normalized. Also, some kind of data normalization might be required to prevent numerical issues like accuracy loss from arithmetic mistakes. Attributes with large starting ranges are likely to dominate an upward descent after first outweighing characteristics with lesser beginning ranges. Since feature space normalization is not applied to the input vectors outwardly, it may be better understood as a kernel perception of preparation than as a specific kind of preprocessing. For instance, in some elements of detection of intrusions statistics, the highest and lowest points in typical and assault are varied by between nine and ten times. Put another way, normalization is a unique kernel mapping method that makes computations easier by converting the data onto a useful plane. The complicated normalization procedure would require an extended period to process because of the massive volume of data points. The chosen Min-Max normalization method is efficient and rapid.

The real information m is transformed linear into the necessary interval max_n, min_n by applying Min-Max Normalization in Eq. (3).

$$m = min_n + (max_n - min_n) * \left(\frac{m - min_u}{max_u - min_u}\right) \tag{3}$$

One benefit of the approach is that it accurately preserves the relationships between the locations in the information. There is no possibility that it could in any way skew the statistics.

C. *Feature Extraction using LDP Method*

When taking into account variations in age, different orientations, and sizes, illuminating effects, and pose variations, regional characteristics outperform global features. Consequently, for AIFR, we have suggested a texture local description. The highly prejudiced local characteristic from the facial components is found by using the suggested descriptor. The variation in the pattern generates a histogram by computing the particular region's pixel differential from the triplet's layout and the relationship among the dual-directional designs. The suggested extracting features method uses the suggested descriptors' local differential pattern [21].

1) Local difference pattern: The appearance and shape of a face evolve with age, making facial recognition harder. To identify a similarity feature that is robust against intra-class variation, research have, consequently, calculated the disparity among pixels of a local region of dimension $a \times a$in an image of dimension $X \times Y$. The suggested feature descriptor covers every local region of a face image.

The set of all the local region variance sequences is represented by Eq. (4),

$$LDP = \left[LDP^{1}_{R^{1}_{u,v}}, LDP^{2}_{R^{2}_{u,v}}, LDP^{3}_{R^{3}_{u,v}}, \ldots..LDP^{Y_i \times Y_j}_{R^{1}_{u,v}}\right] \quad (4)$$

where, u = 1, 2, 3., X − a +1 and v = 1, 2, 3., Y− a + 1.

For a single local region of $LDP^{1}_{uv}, LDP^{1}_{R^{1}_{u,v}}$is calculated. The pixels $A_{u,v}$ in the local region *LR1 i, j* have intensity values$U_{u,v}$, where *∀u,v = 1, 2, 3, …p.*

$$LR^{1d}_{1,v} = \left[A^{1}_{1,v}, A^{2}_{1,v}, \ldots.A^{1}_{1,v}, ..A^{1*a}_{1,v}, A^{p+1}_{1,v}, A^{p+1}_{2,v}, \ldots..A^{a*a}_{a,v}\right] \quad (5)$$

All of the local region pixel intensity values have been organised in three different formatsrow-wise, column-wise, and orderedto compute the local difference pattern of a face image. The calculation of the difference structure is then performed by transposing the resulting data. The pixels in Eq. (5) have been arranged row-wise and assigned to the 1-dimensional array$A^{1*a}_{1,v}$. The intensities of the correlating pixels are numbered as I and organised in the identical order, with an increasing order of magnitude as the superscript as mentioned in Eq. (6).

$$LR^{1^{1d}}_{u,v} = [U^{1}_{u.v}, U^{2}_{u.v}, U^{3}_{u.v} \ldots.U^{a*a}_{u.v}] \quad (6)$$

where, *u,v = 1,2,…a*. The subsequent Eq. (7) is used to compute the row-wise, column-wise, and ordered pattern variance of the intensity of the pixels of local area$k = a \times a$.

$$\gamma^{k}_{u,v+1} = \left[\gamma^{k-1}_{u,v+1} - \gamma^{k-1}_{u,v}\right] \qquad \forall u = 1, \forall v = k - 1 \quad (7)$$

Ultimately, the aforementioned formulas are used for calculating the k^{th}difference pattern. Finding the sum of the directed difference pattern, row-wise difference pattern, and column-wise difference pattern represented as $\gamma^{k}_{u,v}$ will yield the final absolute value of a local region.

The distinction arrangement in a single local region is calculated using the above equation; similarly, all variance patterns are obtained to form the final LDP feature vector, which is used to find the histogram. Eq. (8) displays the difference pattern's final feature vector:

$$LDP = \left[LDP^{1}_{LR^{1}_{u,v}}, LDP^{2}_{LR^{2}_{u,v}}, LDP^{3}_{LR^{3}_{u,v}}, \ldots..LDP^{v}_{LR^{1}_{u,v}}\right] \quad (8)$$

A feature vector Local Difference Pattern (LDP) of a face image contains all of the computed values. To generate a unique code for every neighbourhood, the process entails thresholding the intensity differences and assigning binary values. The resulting patterns draw attention to differences in intensity by highlighting textural details such as corners or edges. For tasks like texture analysis, object detection, and facial recognition, LDP is especially helpful because it efficiently encodes local information that can be useful for differentiating between various regions in an image. All things considered; Local Difference Pattern is a reliable technique for obtaining discriminative features that can improve the efficiency of a range of computer vision applications.

2) Spectral features: The present investigation incorporates MFCC, LTAS, and three spectral characteristics formants. Formants are an illustration of the resonant that occurs in the vocal cords at the peak of high intensity. Since formants change with emotion, they are often used in speech emotion recognition. Then measure the MFCC from a quadratic Me-scale to investigate the significance of low-frequency variables in comparison to high-frequency elements. Since they are highly responsive to variations in sounds at lower ranges, they are often used in voice and recognition of speech mechanisms as they mimic the way people's hearing systems adjust for tone and the exponential signal power ratio of vocal parts of speech signals. Furthermore, Long-Term Average Spectrum (LTAS) has less computational demand than MFCC. The final three formants, the overall mean of the LTAS, the average of the 12 MFCCs, and the ranges, greatest, and lowest constitute the characteristics of segments [22]. Fig. 2 shows the workflow of Mel-frequency cepstral coefficients (MFCC).

The feature-extracting method known as MFCC collects both non-linear as well as linear features, which are necessary for speaker identification. The frequency spectrum employed by the MFCC adjusts the frequency exponentially for frequencies above and below 1 kHz. With MFCC, the crucial component that constitutes sound transmissions may be recorded. The acronym for complex cepstral coefficients is MFCC. Since the MFCC provides both time and frequency information about the signal, it is more beneficial for feature extraction. MFCCs have found widespread application in voice recognition due to their ability to effectively handle dynamic features of the audio data while capturing all non-linear and linearity qualities. Since the sounds have both non-linear as well as linear properties, MFCC is a useful technique for feature extraction. MFCC is a particularly often employed feature in contemporary speech verification/identification methods, according to a number of studies. These factors make MFCC a popular choice for speech recognition systems:

- The discrete cosine transformation (DCT) effectively makes the cepstral features orthogonal.
- Noise from stationary channels is eliminated by subtracting the cepstral average.
- MFCC is less vulnerable to additive noise than other feature extraction techniques like linear prediction of cepstral coefficients.

The following are the methods that MFCC employs to extract features: Initially signal pre-processing is applied to a voice message. It applies pre-emphasis filtering to equalize the exact dimensions. A Hamming is the Window that has been attached to each block to mitigate the edge impacts caused by the window cutting. A discrete cosine transformation is applied to the processed signal, and it is then softened using a series of triangle filters separated along a Me1 Scale.

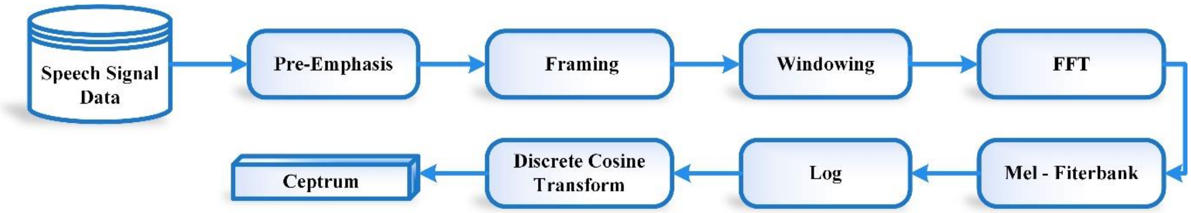


Fig. 2. MFCC workflow.

In audio and voice audio processing, MFCCs are a frequently used feature extraction method. The Fast Fourier Transform (FFT) is used to transform the data into the domain of frequencies, and filter bank processing is then applied to replicate human aural perception. After applying logarithmic compression to the filter bank energies, the MFCCs which capture crucial spectral properties for tasks like sound classification and speech recognition are obtained by the discrete cosine transform (DCT).

D. Feature Selection and Dimensionality Reduction

The purpose of selecting features is to ascertain the characteristics' relative relevance. Following the process of extraction aspects of the input voice signal, only significant characteristics are selected, with the other features eliminated. The characteristics' relevance is calculated at this stage. In this paper, they implement the feature selection procedure using correlational evaluation. In addition, they employ a linear discriminatory analysis approach for feature reducing dimensionality called the Fisher criteria. Initially, Euclidean distance investigation, partial correlation estimation, and vicariate correlation assessment are used to choose the features. The final features with decreased dimensions are then obtained by applying the Fisher criteria to the consequent aspects that were acquired [22].

1) Correlation analysis: The Euclidean distance analysis is completed at this point, and all of the characteristics are grouped together. The correlation between the attributes in each category is then determined by using a modified study of correlation on each group. The final feature set is then determined by evaluating the resulting features for Spearman rank correlation (SRC) assessment. Following the start of the correlation investigation, the resulting emotional traits become increasingly apparent.

2) Euclidean distance analysis: An individual's emotions may be described by a variety of qualities, but it might be difficult to apply them concurrently when conducting emotion recognition. As such, it is crucial to determine which traits are fundamental and have a significant impact on management and emotions. Each characteristic is first examined to see how it relates to other characteristics; those characteristics are then categorised according to the results of this analysis. In this case, the features are grouped using distance analysis. The Euclidean distance shows the true distance between two points in an aspect set of n dimensions. Eq. (9) expresses the Euclidean distance between two points, (X_1,Y_1) and (X_2,Y_2).

$$E = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{9}$$

where, characteristics that have lower *E*scores are clustered into a single group. E indicates the Euclidean distance. X=X_1,X_2, ...,X_n *and* Y=Y_1,Y_2, ..., $Y_n$are locations in a space with n dimensions. Equivalent characteristics are chosen from the closest proximity characteristics.

3) Partial correlation analysis: Since many emotional components share characteristics with a state of feeling, it may be difficult to ascertain how features influence the emotional state. Before examining the link between characteristics and associated feelings, it is necessary to exclude or regulate the features that negatively impact the other aspects. One may refer to this type of study as net correlational analysis or partial correlation assessment. This type of research uses the linear relationship across two features to identify how one affects the other. X= $\{X_1,X_2, ...,X_n\}$ is the set of independent factors; the partial correlation between these variables is calculated using Eq. (10)

$$r = (\rho^{ij})_{n\times n} = \begin{bmatrix} \rho^{11} & \cdots & \rho^{1n} \\ \vdots & \ddots & \vdots \\ \rho^{n1} & \cdots & \rho^{nn} \end{bmatrix} \tag{10}$$

Eq. (11) is used to obtain the inversion for the Matrix mentioned above.

$$r^{-1} = (\lambda^{ij})_{n\times n} = \begin{bmatrix} \lambda^{11} & \cdots & \lambda^{1n} \\ \vdots & \ddots & \vdots \\ \lambda^{n1} & \cdots & \lambda^{nn} \end{bmatrix} \tag{11}$$

Eq. (12) is used to determine the partial relationship among the two variables.

$$Y^{ij} = \frac{-\lambda^{ij}}{\sqrt{\lambda^{ii}}\sqrt{\lambda^{jj}}} \tag{12}$$

The relation between the two separate variables is defined by the coefficient. It subtly illustrates their dependence and the need for selecting or elimination.

4) Nonlinear correlation analysis: There are multiple techniques available when calculating the correlation coefficient between both variables which are a Spearman Rank Correlation (SRC), a Kendall Coefficient of Concordance (KCC), and a Pearson Product Moment Linear Correlation Coefficient (PLCC). However, the characteristics that were taken from all frames inside every time frame are distinct, completely sorted variables, and we used the SRC approach to determine their rank. One kind of index that examines the statistically significant relationship between two variables under a linear function is the SRC Coefficient. The SRC Coefficient is either +1 or - 1 for variables that are strictly monotonic to one another; these variables are referred to be full spearman correlations. Let *X*= $\{X_1, X_2, ..., X_n\}$, and *Y*=$\{Y_1, Y_2, ..., \ Y_n\}$be two variables. Using Eq. (13), the SRC coefficient for each is determined.

$$\rho^s = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \quad (13)$$

In the *X* variable, *Xi*represents i-th characteristic, this is whereas *Yi* represents*i-th* feature in the *Y* variable. The Mean of *X* and *Y* are denoted by $\bar{X}$ and $\bar{Y}$, respectively. A not-parametric correlation value is typically used to calculate the SRC coefficient. When knowing the combined probability distribution of *X* and *Y* samples, the SRC Coefficient calculates precisely the distribution of the two observations. Between *X* and *Y*, the SRC Coefficient is present up through the monotonic connection. The SRC is not the same as the PLCC, which is only dependent on linearity characteristics.

5) Fisher criterion: The multidimensionality of the feature set presents a number of challenges in the development of statistical algorithms for recognised patterns purposes. Low dimensionality methods can provide optimal performance with minimal computing burden. More relevant characteristics are acquired at the feature identification stage, and they undergo a transformation into lower-dimensional space with very little data loss. The loss of data is the main problem of reduction in dimensionality. Therefore, we use the Fisher Criterion, which establishes the concept of linear relation-based reduction in dimensions, to provide an ideal set of attributes with low dimensionality space. Another well-liked technique for reducing complexity is PCA, yet it is unable to separate differentiates between low from very high multidimensional emotional traits. Eq. (14) is used to compute the Fisher Criterion analytically.

$$\lambda^F = \frac{\sigma^B}{\sigma^W} \quad (14)$$

where, σ^W denotes the variation inside the class, σ^B denotes the variance among classes, and λ^F stands for Fisher's rate for characteristics. Eq. (15) definesσ^B.

$$\sigma^B = \sum_{c=1}^{N}(E^c - \bar{E})(E^c - \bar{E})^T \quad (15)$$

where, as specified in Eq. (16),$\bar{E}$ is the average of the whole set of data.

$$\bar{E} = \frac{1}{m}\sum_{i=1}^{m} X_i \quad (16)$$

Moreover,E^c which is specified in Eq. (17), is the sample's average for *ith* Emotions classes.

$$E^c = \frac{1}{N_p}\sum_{X \epsilon E^c} X_i \quad (17)$$

The whole quantity of feelings is denoted by the value *M* in Eq. (16) and the total amount of instances in the speech's emotional signals is denoted by the expression in Eq. (17). Likewise, it has a mathematical definition found in Eq. (18).

$$\sigma^w = \sum_{c=1}^{N}\sum_{i=1}^{N_p}(X_i - E^c)(X_i - E^c)^T \quad (18)$$

They next used decrease in dimensionality to the shape of the distribution vector, σ^w in order to eliminate extraneous features while keeping the crucial data intact.

E. SVM, Decision Tree Classifier

The total amount of characteristics defines *N*, which is utilized to find the hyperplane in the space with *N* dimensions using SVM. The hyperplane facilitates the information point classification procedure. The greatest distance on the plane among the different categories should be used. Non-linear (RBF) kernels are employed in this study to classify support vectors. Using Eq. (4), the kernel aids in obtaining the hyperplane for identifying various classes in Eq. (19):

$$K(X^i, X^j) = e^{-\gamma\|X^i - X^j\|^2} \quad (19)$$

where, the squared Euclidean spacing among the two input information vectors, X^i and $X^j$is represented as $\left\|X^i - X^j\right\|^2$. The incorrect classification rate, C, for the study described in this article, has been fixed at 2. The percentage of inaccurate classifications made by the model that was trained is known as the misperception rate. In order to get the most significant margins between the classes and to signal a smaller erroneous bound, a minimum amount was used [23]. Fig. 3 shows the structure of SVM.

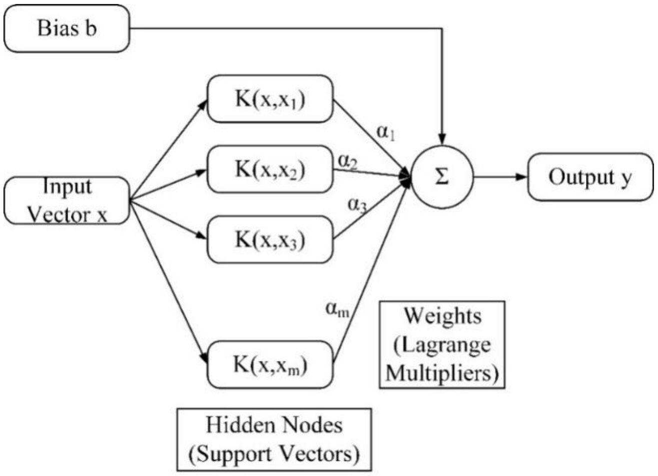


Fig. 3. Structure of SVM.

An optimal splitting of the input characteristics is generated to choose the nodes of a tree that makes up the DT classifier. The greatest information gain (IG) is produced by the separation of the data itself and the tree roots. The tree pruning had been configured with a maximum cutting depth of 8 in order to prevent overfitting of the simulation. Obtaining the parental node's *IG* (20) was applied.

$$IG(D^p) = ID^p - \frac{N_{left}}{N^p} ID_{left} - \frac{N_{right}}{N^p} ID_{right} \quad (20)$$

where, an array comprising the parent, left, and right datasets is represented by D^p, D_{left} , and D_{right}. In this investigation, *I*, the value of entropy, was employed to determine the separated entropy criterion's efficiency. Eq. (21) was used to compute the entropy.

$$I = -\sum_i p^i . log_2 p^i \quad (21)$$

The value that is targeted i'sprobability is represented byp^i. Eq. (22) was utilized in addition to determining the categorization error.

$$DTmodel^{error} = 1 - \max(p^i) \tag{22}$$

F. Self-Similarity Distance Matrix (SSDM)

They represent every movie as a collection filled with regional SSM descriptors H and use recently effective bags-of-features techniques to identify emotion events. Next, they train and categorize examples that represent action classes using SVM.

They create visual text histograms and utilize them as a source of inputs for training using SVM and classifications. Using 10,000 local SSM descriptors (h) divided into k = 1000 clustered from the training collection, a visual language is created. The graphical representation of visual phrases is then calculated for every image in the sequence, and every characteristic is then paired with the vocabulary word that is the closest to it (although they utilize the Euclidean distances). They use the χ 2 kernels for training, not linear SVMs, and use a one-versus-all strategy for the classification of multiple classes [24].

They give n-fold cross-validation findings for each of the recognition investigations in the following section and ensure that a single person's behaviours are not seen in both the training and test sets at the same time.

G. Emotion Recognition Fusion

The methods mentioned above deal with single-frame and 400 ms sound segment prediction. The frame-based forecast was converted to a video-based projection so that the outcomes could be compared with human users. Similarly, to how the audio recordings are divided into manageable chunks, the outcomes were combined to provide one recorded prediction. The following procedure was followed in order to make a single prediction: Each of the six scores for audio and video predictions relates to the expected accuracy for a certain class. Since all probabilities add up to one, the label that is predicted is represented by the value with the greatest sum [25]. To obtain the final forecasting, the individual probabilities are added together and normalized. To get a single forecast for an instance file (audio + video) in a similar fashion, the earlier indicated technique is required. Since the six probabilities form the basis of the audio- and video-predicted labels, these can be easily compared, a process known as decision-level fusion.

V. Result and Discussion

The proposed method's performance evaluation shows notable improvements in the accuracy of emotion recognition. Based on thorough experimental validation using the SAVEE, RAVDESS and RML datasets, the combined strategy shows significant gains over current techniques. A comprehensive feature set that enhances classification is demonstrated by the combination of speech attributes and Weighted Edge Local Directional Patterns for facial pattern extraction.

Metrics for emotion recognition accuracy are displayed in the Table I for a range of emotional categories in three different datasets: SAVEE, RAVDESS, and RML. The reliability for the method of classification is further highlighted by the use of Decision Tree and Support Vector Machine algorithms. The approach gains a valuable dimension with the addition of the novel's Audio-Visual Descriptor, which focuses upon facial alignment as well as selection. This leads to an improvement in emotion recognition performance. The cohesive integration of speech characteristics, multimedia descriptors, and facial patterns enhances the method's adaptability and resilience in identifying feelings within a variety of methods. Based on the percentage accuracy of detecting seven different emotions angry, calm, disgusted, fearful, happy, sad, and surprised each dataset is evaluated. In addition, the SAVEE dataset scored 75% accuracy within the angry category, the RAVDESS dataset obtained 90% accuracy, and the RML dataset showed a high accuracy of 93% accuracy. To identify the remaining emotions within every dataset, the table also gives accuracy percentages. These metrics are useful for comparing how well emotion detection models trained regarding the different datasets perform, showing the different levels of success in correctly classifying different emotional states.

Fig. 4 shows the performance of emotion recognition using accuracy measures for different emotions across three different datasets: SAVEE, RAVDESS, and RML. Particularly, the RML dataset performs well in the angry category with an accuracy of 93%, outperforming both SAVEE (75%) and RAVDESS (90%). The SAVEE dataset is better at identifying Fearful emotions (71%) than the RAVDESS dataset is at identifying Calm emotions (95%). Accuracy in the happy category is comparatively similar, using RAVDESS and RML receiving scores of 86% and 88%, correspondingly. The RML dataset consistently scores well within the disgusted and surprised categories, achieving accuracies of 87% and 88%. The significance of choosing relevant datasets over training models customized to particular emotional states is highlighted by these results, which highlight the dataset-specific details in emotion recognition.

TABLE I. PERFORMANCE OF THE DATASET

Dataset	SAVEE	RAVDESS	RML
Metrics			
Angry	75	90	93
Calm	73	95	90
Disgusted	70	88	87
Fearful	71	88	84
Happy	69	86	88
Sad	70	83	85
Surprised	68	87	88

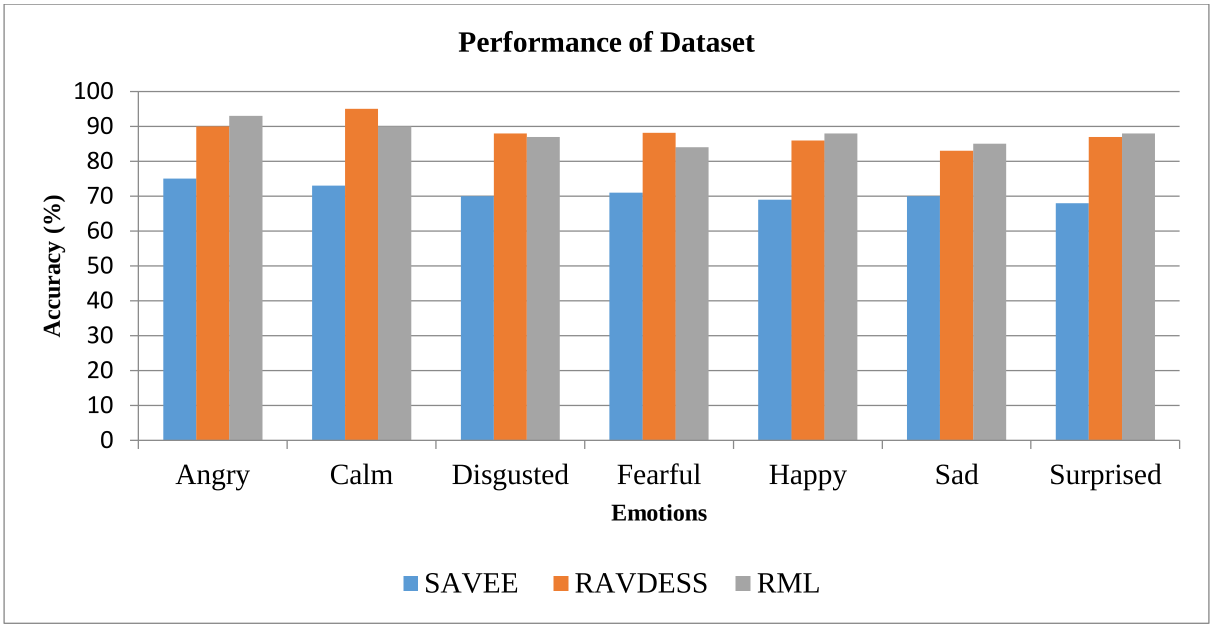


Fig. 4. Graphical representation for performance of the dataset

CNN-LSTM, CNN with Class Activation Mapping, Transformer with Self-attention, the suggested SVM and Decision Tree approach, and other classification techniques are all thoroughly compared in Table II that is presented. With an astounding accuracy of 98.2%, the suggested SVM and Decision Tree technique notably beats its competitors, demonstrating its efficacy in correctly classifying data. The robustness and reliability of the suggested strategy are shown by the precision, recall, and F1-Measure scores of 98.6%, 97.9%, and 98.3%, respectively. With improvements in classification accuracy over the most advanced techniques, this higher performance establishes the suggested SVM plus Decision Tree approach as a highly competitive solution and makes it a viable option for applications needing accurate and dependable classification.

TABLE II. PERFORMANCE COMPARISON WITH EXISTING AND PROPOSED METHOD

Classification method	Accuracy of Classification Method	Precision	Recall	F1-Measure
CNN-LSTM [26]	80%	83.2%	72%	85%
CNN and Class Activation Mapping [27]	90%	93%	90.5%	88%
Transformer and Self-attention [28]	97%	97.8%	98.1%	98%
Proposed SVM and Decision Tree	98.2%	98.6%	97.9%	98.3%

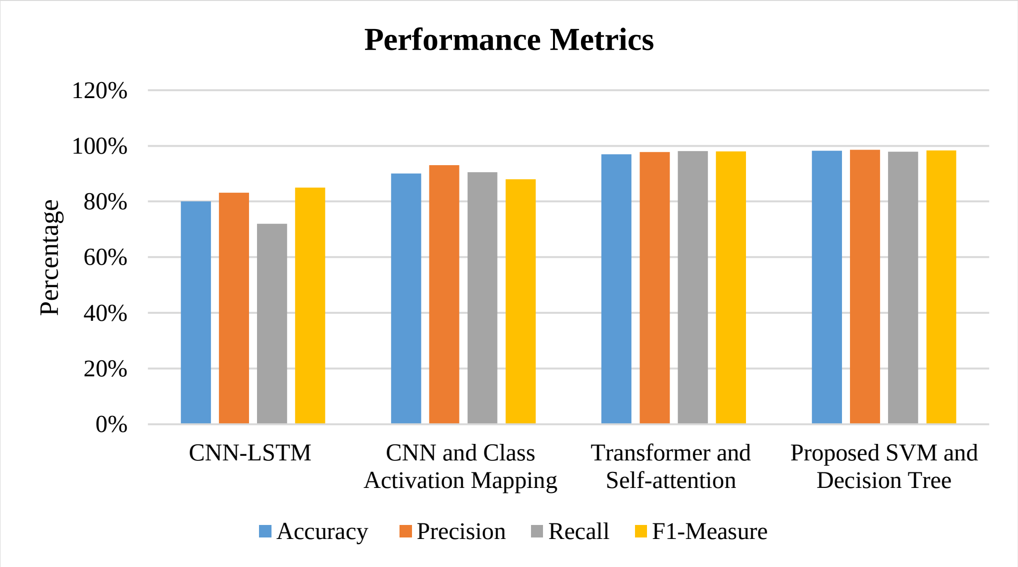


Fig. 5. Graphical representation of performance comparison with existing and proposed method.

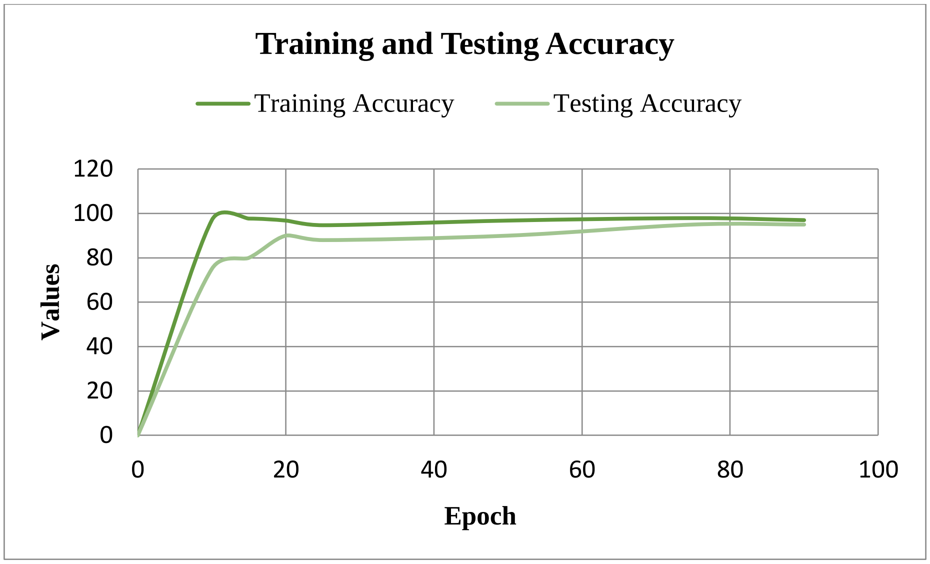


Fig. 6. Graphical depiction for training and testing accuracy of proposed SVM and proposed model.

The performance measures of several classification techniques are shown in the Fig. 5 in terms regarding percentage accuracy. With an accuracy of 80%, the CNN-LSTM model demonstrated its ability to combine long short-term memory (LSTM) networks and convolutional neural networks (CNNs) for classification tasks. A higher accuracy of 90% was shown by the CNN and Class Activation Mapping approach, demonstrating the efficiency of convolutional neural networks enhanced about class activation mapping over localization. With an astounding accuracy of 97%, the Transformer and Self-attention framework demonstrated the effectiveness of self-attention mechanisms in identifying complex patterns. Support vector machines (SVMs) and decision trees work well together to provide accurate classification in the given context. This is demonstrated by the proposed SVM and Decision Tree ensemble, which performed better than other approaches and achieved an impressive accuracy of 98.2%.

Fig. 6 shows the training accuracy graph, within the framework of the proposed SVM and decision tree model, shows how well the model classifies emotions using the training dataset throughout its training iterations, or epochs. This graph shows how the model is learning and if it is becoming more accurate or if the training data may be overfitting. However, the effectiveness of the model on a separate, untested dataset that was not used for training is depicted on the testing accuracy graph. This graph sheds light on how well the model can use its knowledge of emotion recognition to situations outside of the training set. A high testing accuracy shows the model's competence in consistently identifying emotions in a variety of real-world scenarios, and those is essential for improving emotion recognition. A substantial training accuracy suggests that the framework successfully acquired from the training data.

The suggestedSVM and decision tree model demonstrates how the model's loss function changes over the training and evaluation stages in a graphical representation for training and testing loss is shown in Fig. 7. The training loss graph illustrates where the loss, a measure of the discrepancy between expected and actual values, varies throughout the course of the model's training epochs or iterations. The framework learns from and adjusting to the training data when there is a decreasing training loss.The model repeatedly adjusts its variables during training epochs, and the evolution of the loss values which quantify the difference among the predictions made by the model and the actual target values is shown in the training loss curve. The training loss shows that the model is becoming better at fitting the training data. On the other hand, the testing loss curve shows how well the model performs on a test dataset that has not yet been seen, providing information about how well it can generalize and generate correct predictions in practical situations. When faced with new, untested data, the model's capacity to generalize well and provide a lower error rate is demonstrated by the testing loss value that decreases.

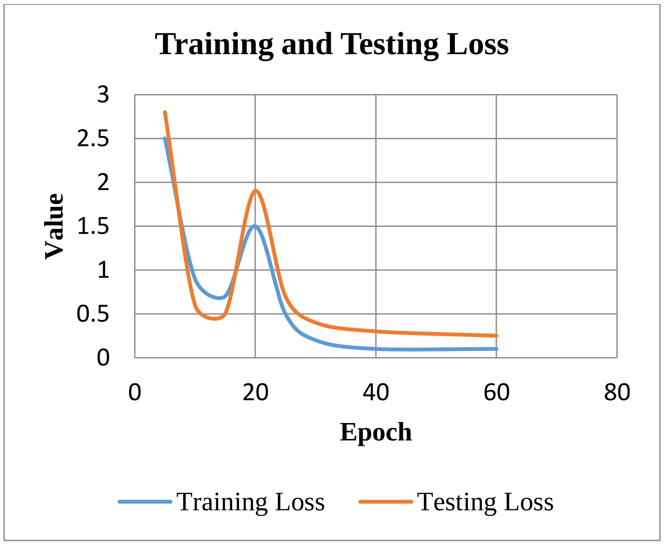


Fig. 7. Graphical depiction for training and testing loss of proposed SVM and decision tree model.

A. Discussion

The outcomes demonstrate the significant progress made by the suggested emotion recognition approach, particularly in contrast to other methods now in use. Using three different datasets, including SAVEE, RAVDESS, and RML, the method's performance is methodically assessed over a range of emotional categories. The increased accuracy of the method's emotion recognition can be attributed to the addition of the Audio-Visual Descriptor, which highlights face alignment and selection. The thorough measurements offer a thorough grasp of how well the approach recognizes seven distinct emotions throughout the datasets. the performance across datasets and emotions, highlighting the subtleties unique to each dataset in the identification of emotions. The suggested SVM and Decision Tree method's exceptional accuracy of 98.2% can be seen when compared to advanced techniques[26] [27] [28]. Its robustness and reliability are further highlighted by the precision, recall, and F1-Measure scores. The suggested approach performs better than CNN-LSTM, CNN with Class Activation Mapping, and Transformer with Self-attention, as demonstrated by the graphical representations. In addition to performing better than existing techniques, the suggested strategy attains a greater accuracy of 98.2%.

The accuracy graphs for training and testing provide information on how the model learns and how well it generalizes to new data. A high testing accuracy shows that the model is useful outside of the training set and demonstrates its competency in real-world circumstances. In addition, the model's learning dynamics are demonstrated by the training and testing loss curves, where lowering loss values denote better generalization and data fitting. The suggested technique for identifying emotions uses a combination of speech characteristics and Weighted Edge Local Directional Patterns, and it shows excellent accuracy and resilience in a range of datasets and emotional classifications. Because of its increased versatility, the Audio-Visual Descriptor presents a viable option for applications that demand accurate and dependable emotion recognition.

This study has certain limitations, despite the notable advancements and gains shown in facial expression analysis and emotion recognition. The suggested model's dependence on edge detection might provide difficulties in situations with different illumination conditions, which could affect the expression assessment's correctness. The model's efficacy can vary depending on the context, and more research may be necessary before applying it to a wider range of real-world settings. The study may have limited cross-platform compatibility and interoperability with other programming languages or frameworks due to its exclusive usage of Python tools for implementation. Further research and development may be needed in the model's ability to handle complex or delicate emotional expressions.

VI. Conclusion and Future Work

In order to overcome difficulties with emotion classification and facial expression identification, this research offers a fresh and practical paradigm. A thorough method for capturing minor emotional cues is demonstrated by the integration of three different feature sets: prosodic characteristics, a new audio-visual descriptor, and Local Differential Pattern (LDP). Specifically designed to improve face feature extraction, the LDP approach is useful for reducing distortion and unreliability associated with edges in facial images. The proposed Audio-Visual Descriptor uses a Self-Similarity Distance Matrix (SSDM) to give a concise description of emotion while giving priority to key frame selection and facial alignment. Experimentation validation using SAVEE, RAVDESS, and RML datasets demonstrates notable advances over existing approaches, highlighting the usefulness of the suggested framework in correctly categorizing emotions and facial expressions in a range of contexts. There are numerous directions for more research and development. More advancement in classification accuracy may be possible by enhancing the suggested framework by adding deep learning architectures, evaluating alternative machine learning algorithms, and examining sophisticated feature extraction methods. Increasing the size and diversity of datasets included in the experimental validation process would improve our comprehension of the generalizability of the model. For practical applications, addressing real-world issues like varied lighting and dynamic face expressions may be essential. It would also be beneficial to investigate how the framework is implemented in real-time settings and evaluate how well it works in erratic environments. In conclusion, the proposed framework raises the bar for the field of emotion identification technology and provides a strong basis for future research projects. Regarding theoretical ramifications, this study makes a substantial contribution to the field by presenting a novel framework that integrates a variety of feature sets, giving emotion categorization a more sophisticated and reliable method. The limitations of current approaches are addressed by the integration of prosodic features, audio-visual descriptors, and LDP, creating opportunities for the exploration of richer emotional cues. From a practical standpoint, the suggested framework has a benefit in that it can reliably categorize emotions and facial expressions in a range of situations. The model has been shown to outperform existing approaches in a number of real-world applications, including affective computing, mental health assessment, and human-computer interaction. These advantages are especially evident in datasets like SAVEE, RAVDESS, and RML. It is essential to recognize the constraints of this research. The accuracy of expression assessment may be affected by the dependence on edge detection algorithms in situations when illumination conditions are changeable. More research and validation are needed to fully understand how the model works in different real-world contexts and how context-specific it is.The suggested framework's potential can be increased for further studies by incorporating deep learning architectures and investigating different methods. A more thorough grasp of the model's capabilities and constraints will also result from extending experimental validation to bigger datasets and tackling real-world issues. Important objectives for future research include investigating real-time implementation and assessing performance in dynamic situations.The foundation for developing emotion recognition technology is laid by this research; however, further investigation and improvement are necessary to ensure the technology's advancement and usefulness.

REFERENCES

[1] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, “Mobile Phone Sensing Systems: A Survey,” IEEE Commun. Surv. Tutor., vol. 15, no. 1, pp. 402–427, 2013, doi: 10.1109/SURV.2012.031412.00077.

[2] R. E. Jack and P. G. Schyns, “The Human Face as a Dynamic Tool for Social Communication,” Curr. Biol., vol. 25, no. 14, pp. R621–R634, Jul. 2015, doi: 10.1016/j.cub.2015.05.052.

[3] S. R. Livingstone and F. A. Russo, “The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English,” PLOS ONE, vol. 13, no. 5, p. e0196391, May 2018, doi: 10.1371/journal.pone.0196391.

[4] “A review on sentiment analysis and emotion detection from text | SpringerLink.” Accessed: Nov. 23, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s13278-021-00776-6.

[5] F. Noroozi, C. A. Corneanu, D. Kamińska, T. Sapiński, S. Escalera, and G. Anbarjafari, “Survey on Emotional Body Gesture Recognition.” arXiv, Jan. 23, 2018. Accessed: Nov. 23, 2023. [Online]. Available: http://arxiv.org/abs/1801.07481.

[6] “Frontiers | SlimMe, a Chatbot With Artificial Empathy for Personal Weight Management: System Design and Finding.” Accessed: Nov. 23, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnut.2022.870775/full.

[7] F. González Hernández, R. Zatarain Cabada, M. Barron Estrada, and H. Rodriguez Rangel, “Recognition of learning-centered emotions using a convolutional neural network,” J. Intell. Fuzzy Syst., vol. 34, pp. 3325–3336, May 2018, doi: 10.3233/JIFS-169514.

[8] S. Shah, H. Ghomeshi, E. Vakaj, E. Cooper, and S. Fouad, “A review of natural language processing in contact centre automation,” Pattern Anal. Appl., vol. 26, no. 3, pp. 823–846, Aug. 2023, doi: 10.1007/s10044-023-01182-8.

[9] S. Shayaa et al., “Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges,” IEEE Access, vol. 6, pp. 37807–37827, 2018, doi: 10.1109/ACCESS.2018.2851311.

[10] F. H. Wilhelm and P. Grossman, “Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment,” Biol. Psychol., vol. 84, no. 3, pp. 552–569, Jul. 2010, doi: 10.1016/j.biopsycho.2010.01.017.

[11] “Experiences of black and minority ethnic (BME) students in higher education: applying self-determination theory to understand the BME attainment gap: Studies in Higher Education: Vol 46, No 3.” Accessed: Nov. 23, 2023. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03075079.2019.1643305.

[12] “Automatic, Dimensional and Continuous Emotion Recognition | International Journal of Synthetic Emotions.” Accessed: Nov. 23, 2023. [Online]. Available: https://dl.acm.org/doi/abs/10.4018/jse.2010101605.

[13] S. Pearson, “Privacy, Security and Trust in Cloud Computing,” in Privacy and Security for Cloud Computing, S. Pearson and G. Yee, Eds., in Computer Communications and Networks. , London: Springer, 2013, pp. 3–42. doi: 10.1007/978-1-4471-4189-1_1.

[14] J. Zhang, Z. Yin, P. Chen, and S. Nichele, “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” Inf. Fusion, vol. 59, pp. 103–126, 2020.

[15] K. Prasada Rao, M. V. P. Chandra Sekhara Rao, and N. Hemanth Chowdary, “An integrated approach to emotion recognition and gender classification,” J. Vis. Commun. Image Represent., vol. 60, pp. 339–345, Apr. 2019, doi: 10.1016/j.jvcir.2019.03.002.

[16] A. Alreshidi and M. Ullah, “Facial Emotion Recognition Using Hybrid Features,” Informatics, vol. 7, no. 1, Art. no. 1, Mar. 2020, doi: 10.3390/informatics7010006.

[17] X. Ben et al., “Video-based facial micro-expression analysis: A survey of datasets, features and algorithms,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 9, pp. 5826–5846, 2021.

[18] Z. Yao, Z. Wang, W. Liu, Y. Liu, and J. Pan, “Speech emotion recognition using fusion of three multi-task learning-based classifiers: HSF-DNN, MS-CNN and LLD-RNN,” Speech Commun., vol. 120, pp. 11–19, Jun. 2020, doi: 10.1016/j.specom.2020.03.005.

[19] E. Ghaleb, M. Popa, and S. Asteriadis, “Metric Learning Based Multimodal Audio-visual Emotion Recognition,” IEEE Multimed., pp. 1–1, 2019, doi: 10.1109/MMUL.2019.2960219.

[20] O. ATİLA and A. ŞENGÜR, “Automatic Speech Emotion Recognition Using Machine Learning and Iterative Neighborhood Component Analysis,” 2021.

[21] R. K. Tripathi and A. S. Jalal, “Novel local feature extraction for age invariant face recognition,” Expert Syst. Appl., vol. 175, p. 114786, 2021.

[22] K. Ramyasree and C. S. Kumar, “Multi-Attribute Feature Extraction and Selection for Emotion Recognition from Speech Through Machine Learning,” Trait. Signal, vol. 40, no. 1, p. 265, 2023.

[23] C. M. T. Khan, N. A. Ab Aziz, J. E. Raja, S. W. B. Nawawi, and P. Rani, “Evaluation of machine learning algorithms for emotions recognition using electrocardiogram,” Emerg. Sci. J., vol. 7, no. 1, pp. 147–161, 2022.

[24] I. N. Junejo, E. Dexter, I. Laptev, and P. Pérez, “Cross-view action recognition from temporal self-similarities,” in Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II 10, Springer, 2008, pp. 293–306.

[25] E. Avots, T. Sapiński, M. Bachmann, and D. Kamińska, “Audiovisual emotion recognition in wild,” Mach. Vis. Appl., vol. 30, no. 5, pp. 975–985, 2019.

[26] E. Ryumina, D. Dresvyanskiy, and A. Karpov, “In search of a robust facial expressions recognition model: A large-scale visual cross-corpus study,” Neurocomputing, vol. 514, pp. 435–450, 2022.

[27] C. Raffel et al., “Exploring the limits of transfer learning with a unified text-to-text transformer,” J. Mach. Learn. Res., vol. 21, no. 1, pp. 5485–5551, 2020.

[28] Y. Zhang, C. Wang, X. Ling, and W. Deng, “Learn from all: Erasing attention consistency for noisy label facial expression recognition,” in European Conference on Computer Vision, Springer, 2022, pp. 418–434.

Research on Efficient CNN Acceleration Through Mixed Precision Quantization: A Comprehensive Methodology

Yizhi He[1], Wenlong Liu[2], Muhammad Tahir[3], Zhao Li[4*], Shaoshuang Zhang[5], Hussain Bux Amur[6]
School of Computer Science and Technology, Shandong University of Technology, Zibo 255049 China[1, 2, 4, 5]
Department of Computer Science, Mohammad Ali Jinnah University, Block 6, P.E.C.H.S, Karachi, 75400, Pakistan[3, 6]

Abstract—**To overcome challenges associated with deploying Convolutional Neural Networks (CNNs) on edge computing devices with limited memory and computing resources, we propose a mixed-precision CNN calculation method on a Field Programmable Gate Array (FPGA). This approach involves a collaborative design encompassing both software and hardware aspects. Initially, we devised a CNN quantization method tailored for the fixed-point operation characteristics of FPGA, addressing the computational challenges posed by floating-point parameters. We introduce a bit-width strategy search algorithm that assigns bit-widths to each layer based on CNN loss variation induced by quantization. Through retraining, this strategy mitigates the degradation in CNN inference accuracy. For FPGA acceleration design, we employ a flow processing architecture with multiple Processing Elements (PEs) to support mixed-precision CNNs. Our approach incorporates a folding design method to implement shared PEs between layers, significantly reducing FPGA resource usage. Furthermore, we designed a data reading method, incorporating a register set buffer between memory and processing elements to alleviate issues related to mismatched data reading and computing speeds. Our implementation of the mixed-precision ResNet20 model on the Kintex-7 Eco R2 development board achieves an inference accuracy of 91.68% and a computing speed 4.27 times faster than the Central Processing Unit (CPU) on the CIFAR-10 dataset, with an accuracy drop of only 1.21%. Compared to a unified 16-bit FPGA accelerator design method, our proposed approach demonstrates an 89-fold increase in computing speed while maintaining similar accuracy.**

Keywords—Convolutional Neural Networks (CNNs); edge computing technologies; Field Programmable Gate Array (FPGA) accelerator; mixed precision quantization; loss variation

I. INTRODUCTION

Nowadays, deep learning has brought new development opportunities for Internet of Things (IoT). Among them, CNNs are widely used in many areas such as face recognition, autonomous driving, and unmanned air vehicles for their outstanding performance [1]. CNN consists of two stages, which are training and inference. Usually, training is a one-time off-line process, which is based on computing platform with large computing resources and high power consumption. Inference can be deployed on edge computing devices, in order to obtain shorter processing delay and avoid the impact of communication situation [2].

However, with the development of deep learning, the size of CNNs continues to increase to obtain stronger learning ability, resulting in larger network computations, more parameters, and more complex network structures. At the same time, many edge computing platforms have limited storage and computing resources, restriction on power consumption and latency. Therefore, the use of CNN inference on edge computing devices has become an important challenge in the field of Artificial Intelligence (AI) research.

Currently, many FPGA acceleration methods [3] for CNNs have been proposed. With FPGA's parallelism and flexible configuration, it can be deeply customized for the CNN structure through parallel computing methods [7] to provide accelerated services for deep learning, achieving higher performance and power efficiency.

Chen et al. [10] proposed the DianNao CNN accelerator, which adopted a three-level pipeline architecture consisting of multiplication, addition, and sigmoid functions. By reusing the weights stored in the on-chip memory, it reduced the need for accessing off-chip data, thereby lowering memory access power consumption. The CNN accelerator Eyeriss [11] was proposed, which employed a row stationary dataflow to maximize data reuse in the computation array and minimize memory access. The authors of this article reference [12] proposed the energy-efficient and reconfigurable hybrid neural network processor Thinker. Each computing unit in Thinker supports adaptive computation for different data bit-widths required by neural networks. The maximum operating frequency is 200MHz, and the supported data bit-widths are 8-bit and 16-bit.

Additionally, quantization [13] can be used to reduce model sizes and hardware resource consumption, such as replacing original 32-bit floating-point operations with lower precision fixed-point numbers like 8-bit or 16-bit. Jacob et al. [15] proposed an integer quantization method, which uniformly quantizes both weight and activation to 8-bit. Furthermore, there are ultra low-bit quantization methods, such as ternary quantization [16] which quantifies weights into {-*w*, 0, +*w*}, and even binary quantization neural networks [17], which quantize weights and activation values to 1 or -1.

However, using a unified quantization bit-width in ultra low-bit-width situations would significantly affect CNN performance. A highly effective solution to this problem is through mixed precision quantization [20]. It allows each layer of the CNN model to have different quantization bit-widths. which can greatly preserve the performance. Lin et al. [21] proposed an analytical solution to address the fixed-point quantization problem. It seeks an optimal bit-width allocation strategy across network layers by optimizing the Signal-to-Quantization-Noise Ratio (SQNR). Wang et al. [22] designed a Hardware Aware Quantization (HAQ) algorithm that incorporates inference speed information evaluated by a hardware simulator into the training process, which utilized reinforcement learning to automatically determine quantization strategies. It reduces latency by 1.4-1.95 times and energy consumption by 1.9 times. However, the current methods face challenges in their applicability to FPGA platforms or in terms of high time and space complexity when searching for mixed precision strategies.

In conclusion, if we have an efficient mixed precision search algorithm and can apply the strategies obtained by this algorithm to FPGA platforms; it will be greatly significant for the application of deep learning on AIoT devices. Therefore, we propose a method for implementing a mixed precision CNN model on FPGA, co-designing from software and hardware aspects. The main innovation points are as follow:

1) A quantization method is proposed that is more suitable for FPGA's fixed-point operation characteristics. Combined with a mixed precision strategy search algorithm with the optimization objective of the lowest bit-width for each layer and the constraint of the accuracy of the CNNs achieve mixed precision calculation of CNNs.

2) In terms of FPGA accelerator design, an inter-layer storage and multi-PEs reuse mechanism is proposed based on the streaming processing architecture. And performing folding design between different CNN layers not only retains the flexibility of the streaming processing architecture but also saves hardware resources and improves the computing efficiency of each PE.

3) A new data reading method is designed using a high bit-width data transmission mode to read multiple data in a single cycle, efficiently utilizing bandwidth to transfer data. In addition, a register set is added as a buffer between the Block Random Access Memory (BRAM) and the PE to solve the problem of mismatch between reading and computing speed.

Based on the above, the purpose of this paper is to explore an efficient method for searching mixed precision quantization strategies and implementing mixed precision computation of CNN on the FPGA platform. The goal is to achieve high performance CNN computations within limited computational and storage resources.

This paper is structured as follows: Section II elucidates the mixed precision quantization method and outlines the strategy for bit-width exploration in the context of CNNs. In Section III, the FPGA platform accelerator is detailed, encompassing the overall architecture, parallel processing elements (PE), and an efficient data transfer mechanism. Section IV delves into the experimental results of quantization and assesses the performance of the mixed precision accelerator. Finally, Section V provides a comprehensive conclusion and future work for the paper.

II. Mixed Precision Quantization for CNNs

A. *Design of Quantization Method*

Although many existing quantization methods can reduce the parameter storage of CNNs through encoding and decoding, the quantized parameters are still computed using floating-point numbers [23]. When implemented to the FPGA platform, the fixed-point number used differs in precision from the original floating-point number, which leads to calculation error and causes drop of inference accuracy. Therefore, a quantization method suitable for FPGA has been designed to convert the parameters in the CNN model into fixed-point numbers, and retrain the quantized model to reduce the degradation of its accuracy. The quantization process consists of two steps: first, the floating-point number is transformed into a fixed-point number, which is left shifting, amplified, rounded, and truncated to a *n*-bit fixed-point integer. Then, the fixed-point decimal value is restored to an approximately original value through right shifting.

The process of converting a 32-bit floating-point number *X* to a *n*-bit integer X_{int} is shown as follows:

$$X_{int} = clamp(round(2^{n-l-1} \cdot X), Q_{min}, Q_{max}) \tag{1}$$

Where *l* is the number of bits in the integer part of *X*, Q_{min}=-2^{n-1}， Q_{max}=2^{n-1}-1， *round*() is the rounding function, and *clamp*() is defined as follows:

$$clamp(x,a,b) = \begin{cases} a, & x < a \\ x, & a \le x \le b \\ b, & x > b \end{cases} \tag{2}$$

For example, the process of quantizing a decimal with a floating-point number of 1.253 to an 8-bit fixed-point number (assuming 1 sign bit, 3 bits of integer width, and 4 bits of decimal width) can be shown in Fig. 1.

The floating-point number 1.253 is shifted left by 4 bits (multiplied by 2^4) and then amplified to obtain 20.048. Then, it is rounded and clamped to obtain an integer value of 20. Finally, the quantized integer value is shifted right by 4 bits (divided by 2^4) to obtain the fixed-point value of 1.25. Hence, it is possible to replace 32-bit floating-point numbers with 8-bit fixed-point numbers in order to make them more easily deployable in FPGA.

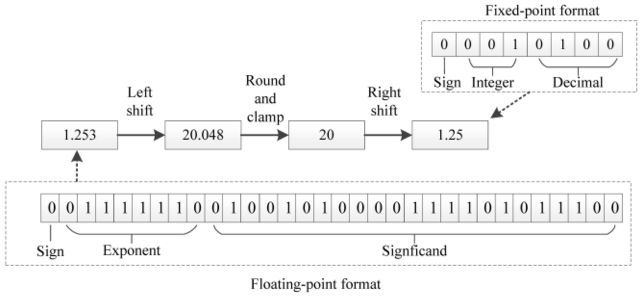


Fig. 1. Quantization Process Diagram.

Suppose the activation value A and weight parameter W in the CNNs are quantized to A_{int} and W_{int} respectively, then the convolutional process can be transformed into,

$$Y = A*W + b = A_{int} * W_{int} \cdot 2^{l_A + l_W + 1 - 2n} + b \tag{3}$$

where b represents bia, l_A and l_W represent the integer bit width of activation value A and weight parameter W, respectively.

Through (3), the floating-point number in the convolution process can be quantized into n-bit integer. After completing the convolution calculation, right shifting can restore the approximate value to the original result. Using shifting and integer operations instead of floating-point arithmetic can reduce the computational resource requirements for floating-point operations, which is easy to implement in FPGA. Moreover, the quantized fixed-point CNN can be retrained, greatly reducing the accuracy drop caused by directly implementing the CNN model to the FPGA platform.

B. Quantization Bit-Width Strategy Search Algorithm

Allocating appropriate quantization bit-widths for each layer in a CNN model is an important challenge in implementing mixed precision operations. Assuming there are 20 convolutional layers in a CNN model, each convolutional layer can be assigned a bit-width value ranging from 1 to 32. There are 32^{20} quantization schemes, and several schemes produce different inference accuracy and occupy different hardware resources. Therefore, selecting a suitable quantization scheme that has high accuracy and is also easy to deploy on FPGA requires an efficient strategy search algorithm. A novel CNN mixed precision quantization method is proposed by using quantization loss variation [25] to adjust the bit-widths of each layer, which can avoid this exponential search space.

1) Calculation of quantization loss variation: The quantization loss variation is an important indicator of bit-width allocation. We found that the loss variation is related to both the first and second derivatives (Hessian matrix) information of each quantization layer and the quantization error. It can be expressed by as shown as follows using Taylor expansion:

$$\Delta L = L(W_Q) - L(W) \approx g(W)^T \Delta W + \frac{1}{2} \Delta W^T H(W) \Delta W \tag{4}$$

g() represents the first derivative of the weight parameter W in the full-precision CNNs, H() represents the second derivative, L() represents the cross-entropy loss function commonly used in CNNs, and W_Q represents the weight parameter of the quantized CNNs. When studying the loss variation brought by second-order information, it is very difficult to directly calculate the relevant values of the Hessian matrix due to the large amount of weight parameters in CNNs. Therefore, the power iteration method can be used to approximate the maximum eigenvalue of the Hessian matrix. For each quantization layer, perturbations can be added in the direction of the corresponding eigenvector of the Hessian matrix as the quantization error [25], and then calculates the corresponding loss variation. Thus, we add perturbations in both gradient direction and Hessian matrix eigenvector direction separately in each quantization layer as the quantization error ΔW, as shown in (5).

$$\Delta W = \begin{cases} \lambda \times g(W) \\ \lambda \times H_i(W) \cdot V_i \end{cases} \tag{5}$$

The λ can be used to adjust the size of perturbation. $H_i(W) \cdot V_i$ represents the eigenvector corresponding to the maximum eigenvalue of the Hessian matrix for the *i-th* quantization layer, which can be calculated by (6) and the power iteration method.

$$\frac{\partial (g_i^T V_i)}{\partial W_i} = \frac{\partial g_i^T}{\partial W_i} V_i + g_i^T \frac{\partial V_i}{\partial W_i} = \frac{\partial g_i^T}{\partial W_i} V_i = H_i V_i \tag{6}$$

V_i is a random vector with the same dimension as the *i-th* quantization layer. The quantization error ΔW in the gradient direction is applied to the selected quantization layer and calculate the perturbed CNN loss variation ΔL_1. Then, the quantization error in the eigenvector direction of the Hessian matrix is applied to this quantization layer and the new loss variation is recalculated as ΔL_2. Finally, the maximum value between ΔL_1 and ΔL_2 is taken as the loss variation caused by this quantization layer.

2) Search method design: To reduce the search space, adjacent quantization layers in the CNNs with the same structural characteristics (such as kernel size, number of channels, and padding method) are merged into quantization blocks. The method described in last section is used to calculate the loss variation of different quantization blocks. High bit-widths are assigned to the quantization blocks that cause large loss variations, while low bit-widths are used for those that cause small loss variations. Then, the model is retrained according to the bit-width allocation strategy. The feedback results from the training are fed back to the policy search module, and the output policy is adjusted based on the feedback until the best bit-width allocation method is found. The specific process is shown in below Algorithm 1.

Algorithm 1: Quantization Bit-Width Policy Search

```
Input：Pre_Model, Dataset, Loss, Blocks, Acc_Set
Output：Bit-Width Policy
Blocks = Sort (Blocks, Loss)
Bit_Width = [16, 15, 14, … 4, 3, 2]
Index = 0
While (Index < length (Bit_width)) do
    New_policy = upgrade_policy (Bit_width [Index], Blocks)
    Model = Quantize (Pre_Model, New_policy)
    for i in range(epoch) do
        Acc = Train (Model, Dataset)
    if (Acc ≥ Acc_Set) then:
        Index = Index + 1
    else:
        Restore (Blocks)
        Choice_next_block (Blocks)
return New_policy
```

In this algorithm, a full precision pretrained model is used as the Pre_Model, and the corresponding Loss variation is used as input. The layer in pretrained model is merged into Blocks. Dataset is used as model training. Then, based on step 1, the quantization blocks are sorted in descending order according to the Loss values they produce. For the quantization block with a larger Loss value, it is quantized first, and the assigned bit-width is not less than that of the quantization block with a smaller Loss value. The quantization bit-width is set from 2 to 16 bits according to step 2, and the Index set by step 3 is used to select the bit-width value. Initially, every quantization block selects 16 bits, as shown in step 5. Next, a new CNN model is quantized based on the strategy generated by step 5, and the model is iteratively trained for epoch times. The highest training result is recorded as Acc, as shown in steps 7 and 8. Then, the highest training result is compared with the required accuracy Acc_Set in step 9 to make a decision. If the requirement is met, the Index value is increased according to step 10, and an attempt is made to reduce the bit-width of the quantization block. If the requirement is not met, the current qutization is restored the last policy, then the next quantization block is selected to adjust the bit-width value according to steps 12, 13. By following the above steps, the appropriate bit-width is selected for each quantization block in sequence.

III. Mixed Precision Accelerator Design

The mixed precision quantization of CNN achieves a balance between accuracy and compression rate by selecting appropriate bit-widths for the weight and activation parameters in each layer [26]. Therefore, in the accelerator, a design method is adopted that uses multiple types of Conv modules to support different bit-width precision operations. This can avoid inefficiency of high-bit-width arithmetic units used by low-bit-width operations. The structure is shown in Fig. 2.

In advance, the calculation method of each layer in CNN model is written into the finite state machine (FSM) in the form of instruction according to the sequence. The input image can be stored in an off-chip memory, and during computation, burst transmission via the AXI4 bus can be used to read the input image from the off-chip memory to the Conv module [27]. The entire computing process only needs to read the input image from off-chip memory once to reduce the high latency and energy consumption caused by off-chip reading and writing operations.

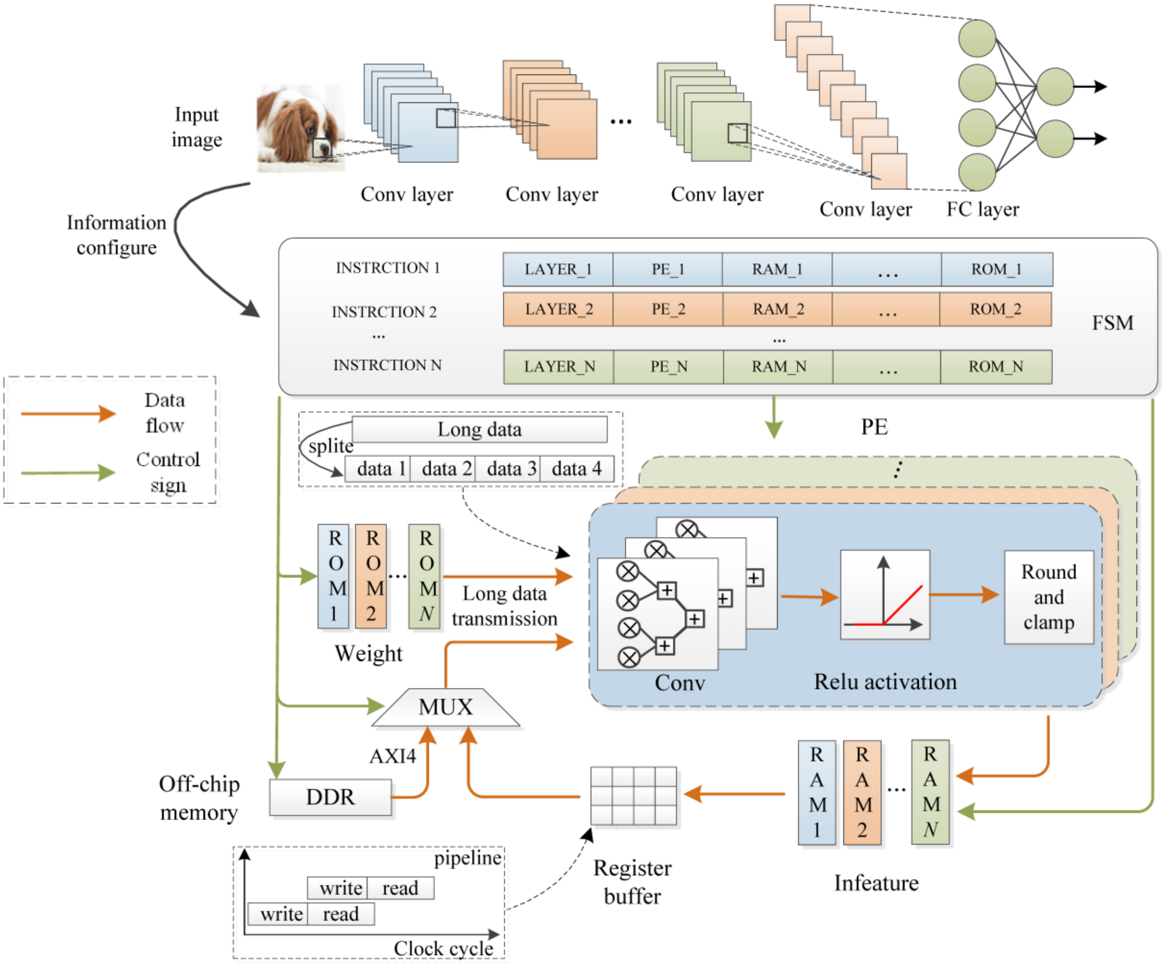


Fig. 2. The mixed precision accelerator design structure.

Then, the FSM sends instruction to control the reading of weight parameters from the Weight Read-Only Memory (ROM) and sends them to the designated PE for convolution, activation, quantization (Round and Clamp). Then the calculation results are cached in the RAM specified by the instruction, and execute subsequent processing according to the next instruction provided by FSM.

For the intermediate buffer data reading and writing operations, their throughput can easily become a bottleneck for the entire accelerator performance due to the limitations of the BRAM reading and writing interfaces. This can cause a problem of mismatch between data calculation speed and reading speed. Therefore, a register set is introduced as a data buffer. The true dual-port RAM and pipeline are simultaneously used, which greatly increase the speed of reading data into PE.

In term of weight parameter storage, the quantized weight parameters of each layer in the CNN are stored in the ROM array (Weight ROM) built of on-chip RAM resources according to the address order. During circuit initialization, the parameters are stored in the on-chip memory in a ROM initialization method. The read width is set to the sum of all weight parameter bit-widths in a single convolution kernel. Data slicing allows for extracting specific bits from a signal by using indices. So the entire long bit-width data can be split and sent sequentially to the Conv module after reading the data. In ROM IP core provided by Xilinx, the maximum read width can reach 4608 bits, which is sufficient to read the weight parameters in a convolutional kernel within one cycle.

The overall processing adopts a streaming processing architecture, and the CNN is folded according to the mixed precision quantization results in order to share PEs and memory resources among different layers. This not only alleviates the problem of large resource consumption of the streaming processing architecture, but also allows the saved resources to be used in each Conv module to improve computation parallelism. Meanwhile, the entire network can still be designed with pipeline processing. Each computing module can serve as a stage pipeline, forming a large-scale pipeline design to reduce computation latency under the condition of multiple input data.

A. PE Design

The PE structure is shown in Fig. 3. It is composed of the Conv module, Relu activation, Round and Clamp module. In the same PE, all multipliers are the same type. For example, the multipliers in Fig. 3 are all $M\times N$ type, where M corresponds to the weight bit-width and N corresponds to the feature data bit-width. The number of multipliers in each group is set according to the convolution kernel size. Assuming the convolution kernel size is $K\times K$, then K^2 multipliers are set as one group. The number of groups S is set to the least common multiple of input channels of all quantization layers using this PE. This can fully utilize all multipliers and improve computation efficiency when calling this PE.

In each convolution module, there are multiple layers of adder trees and a multiplexer. The number of adder operations can be controlled based on the input channel number of the current layer to adapt to different computation modes with different input channel numbers. For example, if the convolution channel S is set to 64 in the PE and the current network layer has 32 input channels, then two output pixels can be obtained through a binary adder tree of depth 5. When computing a network layer with 64 input channels, one output pixel needs to be obtained through a binary adder tree of depth 6. Therefore, we can reuse the PE for different layers by setting a multiplexer to output results at different layers of the adder tree.

In CNN models, to accelerate the convergence speed, prevent problems such as gradient explosion, gradient vanishing, and overfitting, many networks have Batch Normalization (BN) layers [28]. The process of calculating the BN layer with convolutional result Y through (7), can be represented as:

$$Y_{bn}=\gamma\cdot(\frac{Y-\mu}{\sqrt{\sigma^2+\varepsilon}})+\beta=\gamma\cdot(\frac{W*A+b-\mu}{\sqrt{\sigma^2+\varepsilon}})+\beta \tag{7}$$

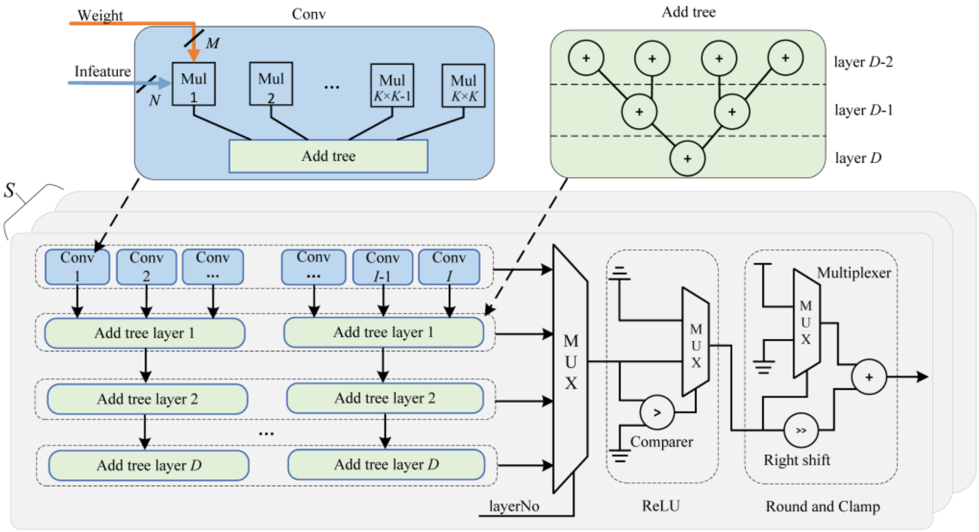


Fig. 3. Structure diagram of the PE unit.

Where *μ* and *σ* represent the mean and standard deviation within a batch, *γ* represents the scaling parameter, *β* represents the offset parameter, and *ε* is a very small constant set to 0.001. However, the additional computation brought by this layer makes it more difficult to implement the CNNs on FPGA platforms. Thus, the convolution, batch normalization and quantization operation are integrated together to solve this problem. Firstly, the process of integrating convolution and BN can be represented as follows:

$$Y = W' * A + b' \tag{8}$$

ωηερε $W' = \dfrac{\gamma W}{\sqrt{\sigma^2 + \varepsilon}}$, $b' = \dfrac{\gamma(b - \mu)}{\sqrt{\sigma^2 + \varepsilon}} + \beta$

Furthermore, the quantization operation (3) can be merged into (8) as follows:

$$Y = W'_{int} * A_{int} \cdot 2^{l_A + l_W - 2n+1} + b' \tag{9}$$

The low bit-width fixed-point weight W_{int} can be obtained directly by using the optimal parameters for the weight *W'* and quantization bit-width *n* and scale factor l_A. All of them can be obtained from algorithm 1. By integrating the three operations, many complex calculations can be completed during quantization training, the trained parameters W'_{int} can be stored directly in the on-chip ROM of FPGA during circuit initialization. So, this significantly reduces lots of calculations on the FPGA.

In addition, an activation layer is often used after the convolution operation to increase the non-linear ability of the CNN. To simplify the design, we also integer the activation layer into PE. The logic structure of the commonly used ReLU activation function is shown in the Fig. 3, which uses a comparer and multiplexer. The comparer compares the input value with zero, and the result controls the multiplexer to output either the input value or 0 as the activation value.

The convolution operation with *M* bits weights and *N* bits input feature values will result in *M+N* bits of convolution results. However, in the next layer of the network, the bit-width for the computation has already been specified by the quantization strategy, and the convolution result needs to be quantized to the specified bit-width for the next layer. Therefore, we perform Round and Clamp operations on the activated convolution result to truncate the length of the data to the specified bit-width as the input feature value for the next layer. In this way, the intermediate calculation results can also be stored in low bit-width BRAM, which saves the storage resources. The Round and Clamp operations consist of a multiplexer and an adder, which determine whether the first bit after the reserved bits of the data is equal to 1. If it is 1, the rounding operation will add 1 to the reserved data result. Otherwise, it will keep the original value.

With the above design, calculating one pixel in the output feature map requires multiplication unit, adder tree, activation, and quantization process. Considering the delay caused by these processes, it is difficult to output one result value in each cycle. Therefore, we perform pipeline to improve computing efficiency, and the specific process is shown in Fig. 4.

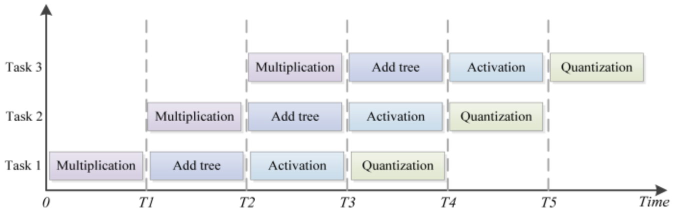


Fig. 4. The PE pipeline.

We assume that each process requires one clock cycle *Δt*, and calculating one output pixel requires *m* processes. According to the above settings, if no pipeline is applied, the number of cycles required to calculate all the pixels in the output feature map can be expressed as:

$$T_i = mH_{out}W_{out}\Delta t \tag{10}$$

W_{out}, H_{out} represent the width and height of the output feature map, which can be calculated based on the convolution process.

$$\begin{cases} H_{out} = \dfrac{H_i - K + 1}{S} \\ W_{out} = \dfrac{W_i - K + 1}{S} \end{cases} \tag{11}$$

H_i, and W_i are the height and width of the input feature map, *K* is the size of the convolution kernel, and *S* represents the convolution stride. With the pipeline, every clock cycle can generate one output pixel value except for the first output pixel point calculation process. The total number of calculation cycles can be reduced to the result expressed in (12). By comparing (10) and (12) , the computing efficiency has got improved significantly.

$$T_i = (H_{out} \cdot W_{out} - 1)\Delta t + m\Delta t \tag{12}$$

B. Folding Design

In one layer of the CNNs, there are many identical structures of channels, and even in many CNNs, there are many layers with the same structure and convolution method. When there are many channels in each layer, full parallel design according to the layer order in FPGA will inevitably use a large amount of calculation and storage resources, which may even exceed the existing resources of the FPGA [29]. Moreover, in CNNs, data needs to be passed in layer order, and it is difficult to parallelize between layers. Therefore, based on the above two calculation characteristics, we fold the original data flow to reduce the number of PE, and map it to hardware using the method of PE reuse. So different network layers can share PE while still keep the original computing efficiency. The process is shown in Fig. 5.

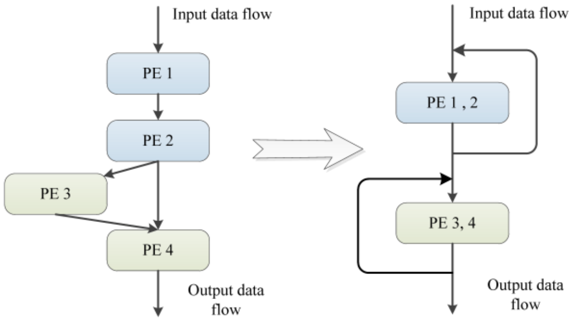


Fig. 5. Streaming folding diagram.

When different layers share the same PE, a control module needs to be introduced because there are differences in weight parameters, convolution strides, and other information between each layer of the CNN. The control module can dynamically switch the computing mode according to the requirements of different layers, such as adjusting the address for loading pre-trained weights, the source of input feature, and the address for storing intermediate calculation results. So, we need to pre-analyze the structural parameters of the selected CNN, compile the layer number, the storage address of the pre-trained weight, the convolution stride, the BRAM position where the intermediate cached result is stored, and other information of each layer into one instruction data and write it into FSM. After FSM outputs instructions, they are divided into segments. The calculation and data flow in FPGA can be uniformly allocated through various instruction segments. The length L of each segment can be macro-defined according according to (13).

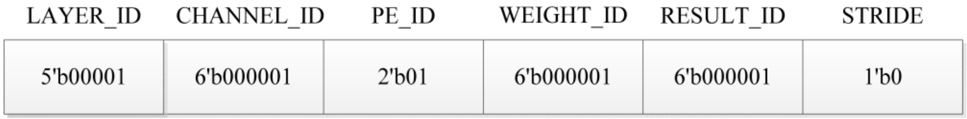
$$L_i = \lceil log_2 N_i \rceil \quad (13)$$

N_i represents the number of species included in the *i-th* segment. If it is assumed to perform calculations on the first channel (64 channels in total) of the first layer (20 layers in total) of the network, the format of this instruction provided by FSM is shown in Fig. 6.

C. Optimization of Data Reading Methods

In the design of mainstream accelerators [31], to avoid high latency caused by accessing off-chip memory, weight parameters and intermediate calculation results are usually stored into on-chip BRAM memory in sequential order directly. However, when reading data, only one or two data can be read per clock according to the corresponding address. After quantization to low-bit-width fixed-point numbers, if the data is still transmitted according to the previous method, it will result in underutilization of bandwidth resources and a mismatch between data reading speed and computation speed. Therefore, a new data reading method is proposed to reorganize low-bit-width data by concatenating multiple low-bit-width data into one long word. At the same time, true dual-port RAM and register set are used in combination with data reuse to improve reading speed.

This method for reading data through 3×3 convolution is explained as follows: When reading the data, it is necessary to first set up a two-dimensional register set as a buffer between RAM and Conv module, as shown in Fig. 7.



LAYER_ID	CHANNEL_ID	PE_ID	WEIGHT_ID	RESULT_ID	STRIDE
5'b00001	6'b000001	2'b01	6'b000001	6'b000001	1'b0

Fig. 6. The instruction format.

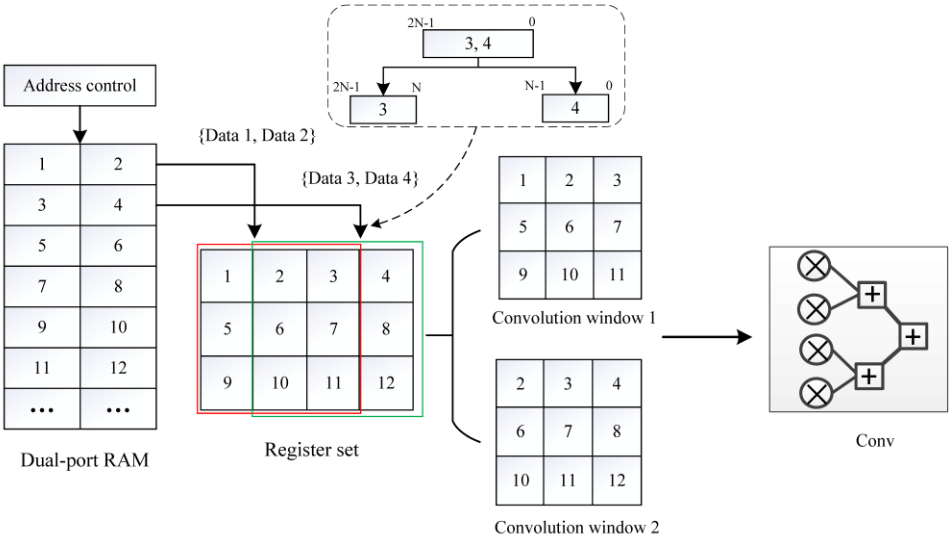


Fig. 7. Buffer structure diagram

Double-port RAM is a type of memory component that has two independent data ports. The write data width of a double-port RAM is set to twice the read data width, *2N* bits. When reading the data, two addresses are set as adjacent numbers, so that four adjacent data can be read and stored into the first row of the two-dimensional register set within one clock cycle. Then, in the next clock cycle, four adjacent data are read according to the address and stored into the second row of the register set. After the register set is full, the data from columns 1 to 3 can be directly read to achieve 3×3 full parallel convolutional calculation. Then, the data from columns 2 to 4 can be read to complete the second convolution, while reusing the second and third columns of data, thus reducing the overall data access. Additionally, when FPGA resources are sufficient, a larger register set and higher reading data width can be set, which not only can realize a higher data reuse ratio, but also can achieve a faster pipeline operation.

IV. Experiments and Result

To verify the effectiveness of the proposed method in this paper, the Resnet20 model was selected for validation on the CIFAR-10 dataset. The deep learning framework PyTorch was used to complete the quantization training experiment and mixed precision strategy search process on a server equipped with an NVIDIA Tesla T4 GPU. The quantized Resnet20 model was designed using the Verilog language on a Kintex 7 Eco R2 development board, and data reports were generated through synthesis and implementation with Vivado 2019.2 to analyze resource utilization and power consumption.

A. Resnet20 Quantization Experiment

The full precision Resnet20 model is iteratively trained on the CIFAR-10 dataset for 300 times. The accuracy of the training result is 92.89%, which was used as the pretrained model for quantization. We set the inference accuracy of the quantized model to not be less than 91.5%. The main network structure of Resnet20 is shown in Fig. 8, which has three types of residual blocks, with the convolution kernel size, channel number, and padding mode being the same in each type. According to this structural feature, we divided the network into five major structural blocks. The first layer is the first structural block, layers 2-7 are the second structural block, layers 8-14 form the third structural blocks, layers 15-21 form the fourth structural block, and the last fully connected layer is the fifth structural block.

Based on the above settings, we search for the quantization strategy in Algorithm 1, the final quantization bit-width results of the Resnet20 are shown in Fig. 9.

The activation and weight bit-widths of the first and fifth structural blocks are both quantized to 8 bits, the weight bit-widths of the second and third structural blocks (layers 2-14) are quantized to 7 bits, and the weight bit-widths in the fourth structural block (layers 15-21) are quantized to 8 bits, with the activation bit-width being 7 bits.

After mixed precision quantization and retraining of Resnet20, the inference accuracy is 91.68%, with only a 1.21% loss compared to the full precision Resnet20.

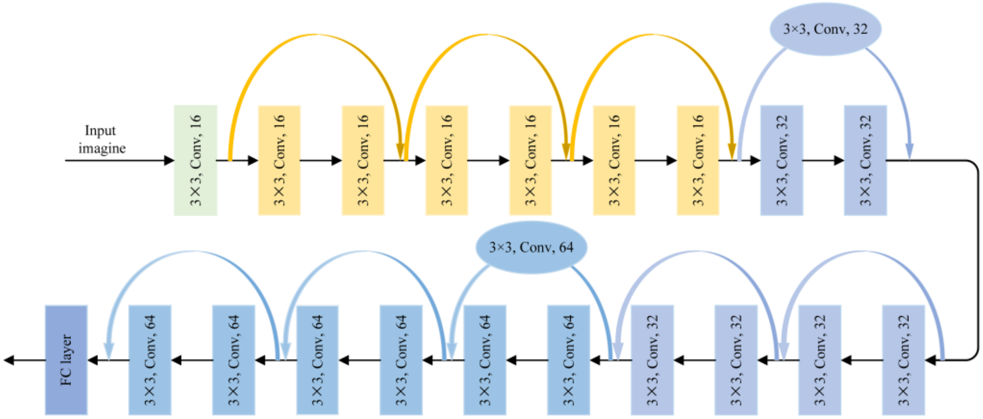


Fig. 8. Partial structure diagram of Resnet20.

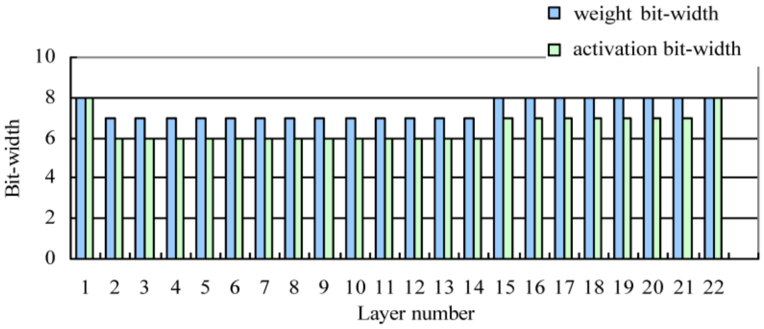


Fig. 9. Quantization bit width strategy for Resnet20.

Table I shows the comparative results of related work. In [33], the method of uniformly quantizing to 8 bits was used, and the accuracy was 90.7%. But in our paper, by reasonably allocating different bit-widths for each quantization layer, the accuracy is still 0.98% higher than their method even with lower bit-widths. In [34], after quantizing the full precision model to 8 bits and directly transplanting it to the FPGA platform, the calculation error caused by replacing the original floating-point numbers with fixed-point numbers resulted in an accuracy loss of 6.92%. By comparing this method, it can be seen that our paper greatly reduces the loss of CNN accuracy by performing quantization training and selecting a more suitable quantization method for the FPGA, making the quantized CNN closer to the full precision one.

B. Accelerator Design Experiment

1) Analysis of FPGA resource usage: Firstly, we tested the resource requirements for different bit-width multipliers in this experiment by using the synthesis tool in Vivado 2019.2, as shown in Table II. It can be seen that it takes 2 Digital Signal Processors (DSPs), 128 Look UpTables (LUTs) and 299 Flip Flops (FFs) to complete 32-bit floating-point multiplication. If DSP is not used, it would require 606 LUTs and 805 FFs to construct a 32-bit floating-point multiplier. By quantizing the multiplication operation to no higher than 8 bits, the resource usage can be reduced significantly. For example, the 7bits×bits multiplier only needs 45 LUTs and 13 FFs. The main chip xc7k325tffg676-2 in the development board used in this experiment has only 840 DSPs, but it has 203800 LUTs. Despite the limited number of DSPs available, additional multipliers are still required to achieve high parallel computations. So both DSPs and LUTs are needed to be used in combination to perform convolution operations in this paper.

Table III displays the Multiply-Accumulate operations (MACs) and resource usage and for each PE. According to the quantization results of Section IV.A in the first block, calculations are performed using 8bits8bits multipliers, therefore, PE 1 is equipped with 432 DSPs of the 8bits×8bits type, allowing for concurrent processing of 48 sets of 3×3 convolutions. The 7bits×6bits multipliers can be used for the calculations in both the 2nd and 3rd structure blocks according to the quantization results. So, these two structure blocks (the 2nd to 14th quantization layers) are designed to be folded and share a single PE (PE 2) with 1152 multipliers, which can compute up to 128 parallel convolutions of 3×3. The PE 3 used in the fourth structure block can use 8bits×7bits multipliers, so the 15th to 21th quantization layers are folded and also designed with 1152 multipliers. In PE 2 and PE 3, all the multipliers are implemented using LUTs to compensate for the limited on-chip DSP resources. In the 22nd layer, which is a fully connected layer, the computation process is carried out in PE 4. With its structure consisting of 64 inputs and 10 outputs, it is designed with 10 DSPs capable of simultaneously performing 10 parallel Multiply-Accumulate operations.

TABLE I. QUANTIZATION EXPERIMENT RESULTS OF RESNET20.

Quantization Method	Weight Bit-Widths	Weight Integer Bit-Widths	Activation Bit-Widths	Activation Integer Bit-Widths	Accuracy (%)
Baseline	32	/	32	/	92.89
[33]	8	/	8	/	90.7
[34]	8	/	8	/	84.81 (91.73)
Ours	7/8	2	6/7/8	4	91.68

TABLE II. COMPARISON OF RESOURCE USAGE FOR DIFFERENT MULTIPLIERS

Multiplier Type (*m*-bit × *n*-bit)	LUTs	FFs	DSPs	Power (W)
32×32 (float)	128	299	2	0.184
32×32 (float)	606	805	0	0.191
16×16 (fixed)	280	32	0	0.192
8×8 (fixed)	71	16	0	0.173
8×8 (fixed)	0	0	1	0.173
8×7 (fixed)	63	15	0	0.172
7×6 (fixed)	45	13	0	0.17

TABLE III. RESOURCE USAGE FOR PE UNITS

Quantized layer number	Convolution kernel size	MACs	PE number	Multiplier type (*m* bits×*n* bits)	LUTs	DSPs	FFs	BRAMs
1	16×3×3×3	884736	1	8×8	7040	432	11008	8
2-7	16×16×3×3	28311552	2	7×6	67768	0	40392	48
8-14	32×16×3×3/32×32×3×3	28311552	2	7×6	67768	0	40392	48
15-21	32×64×3×3/64×64×3×3	28311552	3	8×7	89528	0	49144	96
22	64×10(fully connected layer)	1280	4	8×8	325	10	166	0

TABLE IV. RESOURCE CONSUMPTION FOR RESNET20

Resnet20	Resource usage	Available	Utilization (%)
LUT	186257	203800	91.39
FF	148385	407600	36.4%
DSP	442	445	52.6
BRAM(36Kb)	152	840	34.2

The resource requirements for the entire Resnet20 are shown in Table IV, where the DSP resource utilization rate is 52.6% and LUT resource utilization rate is 91.39%. DSP resources are used for the first quantized layer and the last fully connected layer, while both shared accelerators are constructed entirely using LUTs with low bit-width multipliers to minimize resource usage. Storage areas can be recycled, thereby saving a large amount of BRAM resources.

2) Analysis of the buffer reading and writing: In the Resnet20, all convolution kernel sizes are 3×3, so only one data reading and writing method needs to be designed. This paper sets the register set size to 6×8, and each set can hold 48 data to provide the data required for 24 convolutions. Using the method described in Section 3.3 to read and write data, four data can be written into the register set per cycle, so it will take 12 cycles to fill the entire register set.

To achieve efficient computation, pipeline is added to the writing and reading operations. The specific process is shown in Fig. 10. When reading data from rows 4-6, new data is written into rows 1-3, and the data in rows 4-6 can be used to complete reading and calculation in 6 cycles. During these 6 cycles, the data in rows 1-3 can be updated by writing new data into them. Similarly, when reading data from rows 1-3, new data is written into rows 4-6. This can achieve the ability to read 9 data required for a 3×3 convolution every cycle.

3) FPGA accelerator performance analysis: In order to validate the advantage of FPGA accelerator, we evaluated the computational efficiency of three different platforms: CPU (Intel Core i5-12400), GPU (NVIDIA RTX 2070 SUPER), and FPGA. Table V shows the results of Resnet20 model inference time and energy consumption on CIFAR-10 dataset using three different platforms. Since the CPU and GPU platforms perform better with large batch sizes, we set the batch size to multiple values to obtain their highest performance. The inference time per image was obtained by dividing the total time by the batch size.

The experimental results show that, the computing speed is 4.27 times faster than that of the CPU after quantization, layer fusion, and parallel computation using FPGA, and the required power only needs 5% of the CPU's power consumption. Moreover, because we design a large number of parallel PEs while also deeply customizing the CNN structure, our work achieves similar computational speed under a power consumption of only 6.2% of the GPU platform. Since our design retains the characteristics of the streaming processing architecture, it can still adopt pipeline design when applied to the computation of multiple sets of input images, and each PE can serve as a stage pipeline, which can further improve the overall computational performance.

TABLE V. COMPARISON WITH OTHER EXPERIMENTAL PLATFORMS

Platform	Frequency (MHz)	Latency (ms)	Power consumption (W)	Speedup ratio
CPU	2500	2.01	65	1
GPU	1605	0.54	215	3.72
FPGA	100	0.47	13.31	4.27

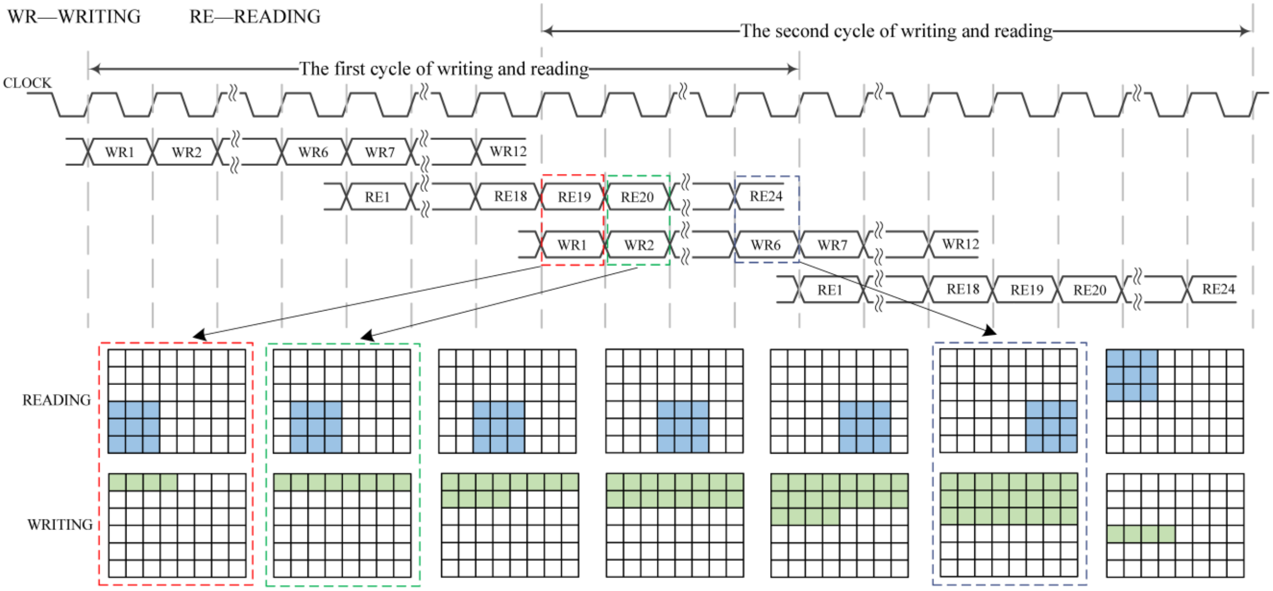


Fig. 10. Reading and writing data pipeline diagram.

Table VI shows the comparative results of other related work. In [35], the FFT method is used to reduce a certain amount of computational complexity, but still uses 16 bits long-data for computation. Due to the high computational complexity brought by long-bit-width data operations, in the case of limited DSP resources, it limits the computational parallelism. But multiple types of PEs are used in our paper which is suitable for different types of low-bit-width data operations to avoid low computational efficiency of a single long-bit-width PE structure. Therefore, compared with this method, our paper shortens the time by 89.3 times and reduces power consumption by 21.8 times. In addition, our paper reasonably allocates bit-widths for each layer of the CNN to improve its accuracy by 0.43%. In [36], they deploy a binary Resnet20 using an Application Specific Integrated Circuit (ASIC), and all network parameters are quantized to 1 bit. This method reduces computational complexity to the minimum via ultra low bit-widths, achieves higher frequency of operations and lower power consumption as shown in Table VI. However, this ultra low bit-width quantization method also produces a large precision error, which greatly affects the performance of the CNN. But the mixed precision quantization method and retraining the quantized model in our paper achieve a higher accuracy rate of 9.88% than that. In terms of computational speed, by efficiently utilizing the development board resources and increasing the parallelism of convolution calculation through multiplication circuits built with LUTs, it has achieved lower computational latency. In [37], the Winograd algorithm is used to reduce the computational load, but 16-bit data operation limits the overall performance, the mix precision low bit-width CNN in our work is more effective.

TABLE VI. PERFORMANCE COMPARISON

Work	Experiment Platform	Frequency (MHz)	Bit-Width (bits)	Latency (ms)	Throughput (GOP/S)	Energy (mJ)	Accuracy
[35]	Zynq 7020	154	16	42	/	137	91.25
[36]	ASIC 65nm	1000	1	0.98	/	3.8	81.8
[37]	ZynqZ7035	150	16	/	43.5	/	/
our	Kintex 7	100	6-8MP	0.47	179	6.26	91.68

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a high-performance CNN design method tailored for edge computing. Employing quantization methods and a strategy search algorithm in the software algorithm mitigated the significant accuracy loss associated with quantizing CNN models. In the FPGA accelerator design, we implemented a reuse structure based on a streaming processing architecture. This involved designing different Processing Elements (PEs) according to the characteristics of the CNN model structure and quantization bit-width. Notably, different network layers could share the same PE, optimizing resource utilization. For efficient data transmission, we adopted a strategy of packing quantized low-bit-width data into long words. This approach fully leveraged high-bandwidth data transfer, utilizing a register set as a buffer and employing a data reuse method to achieve synchronous data reading and computing. The validation of our method using Resnet20 on the CIFAR-10 dataset demonstrated its effectiveness. Comparative analysis with other computational platforms and related works revealed that our CNN accelerator outperformed a unified 16-bit FPGA accelerator design, achieving an 89-fold increase in computing speed with lower power consumption. Specifically, our CNN accelerator exhibited a computing speed 3.72 times faster than the GPU (RTX 2070 SUPER), while consuming only 6.2% of its power. In conclusion, our research presents a novel approach to high-performance CNN design for edge computing, showcasing substantial improvements in computing speed and power efficiency compared to existing methods. As part of future work, we plan to explore further optimizations and scalability of our approach, addressing potential challenges and extending its applicability to broader CNN architectures and datasets.

ACKNOWLEDGMENT

This research was funded by the National key R&D Program of China (2022YFE0107300).

REFERENCES

[1] Q. Jian, P.Y.Zhang and X. J. Wu. “FPGA implementation method for a configurable CNN Co-accelerator,” Journal of Electronics, vol. 47, no. 7, pp. 1525-1531, 2019.

[2] X. Peng. , J. Yu, B. Yao. L. Liu, Y. Peng. “A Review of FPGA-Based Custom Computing Architecture for Convolutional Neural Network Inference,” Chinese Journal of Electronics, vol. 30, no. 1, pp. 1-17, 2021.

[3] K. Guo. “Angel-Eye: A Complete Design Flow for Mapping CNN Onto Embedded FPGA,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 37, no. 1, pp. 35-47, 2018.

[4] Y. Yu, C. Wu, T. Zhao, K Wang, and L. He, “OPU: An FPGA-based overlay processor for convolutional neural networks,” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 28, no. 1, pp. 35–47, 2020.

[5] Y. Wu, L. Kai, Y. Liu, et al. “Progress and trend of deep learning FPGA accelerator,” Chinese Journal of Computers, vol. 42, no. 11, pp. 2461-2480, 2019.

[6] A. Shawahna, Sait. S. M, El-Maleh. A. “FPGA-based accelerators of deep learning networks for learning and classification: a review,” IEEE Access, vol. 7, pp. 7823-7859, 2018.

[7] Z. J. Lin, X. W. Gao, X. P Chen, et al. “Design of high parallel CNN accelerator based on FPGA for AIoT,” The Journal of China Universities of Posts and Telecommunications, vol. 29, no. 05, pp. 1-9, 2022.

[8] Q. Dou, Y. Deng, R. Deng, et al. “Laius: an energy-efficient FPGA CNN accelerator with the support of a fixed-point training framework,” International Journal of Computational Science and Engineering, vol. 21, no. 3, pp. 418-428, 2020.

[9] Y. Ma, Y. Cao, S. Vrudhula, J.S. Seo, “Optimizing Loop Operation and Dataflow in FPGA Acceleration of Deep Convolutional Neural Networks,” In Proceedings of the ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, pp. 45–54, 2017.

[10] T. Chen, Z. Du, N. Sun, “DianNao: A Small-Footprint High-Throughput Accelerator for Ubiquitous Machine-Learning,” SIGARCH Computer Architecture News, vol. 42, no. 1, pp. 269-284, 2014.

[11] Y. Chen, T. Krishna, J. S. Emer, V. Sze. “Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks,” IEEE Journal of Solid-State Circuits, vol. 52, no. 1, pp. 127-138, 2017.

[12] S. Yin, P. Ouyang, S. Tang. “A High Energy Efficient Reconfigurable Hybrid Neural Network Processor for Deep Learning Applications,” IEEE Journal of Solid-State Circuits, vol. 53, no. 4, pp. 968-982, 2018.

[13] X. Ruan, W. Hu, Y. Liu. “Dynamic sparsity and model feature learning enhanced training for convolutional neural network-pruning,” SCIENTIA SINICA Technologica, vol. 52, no. 5, pp. 667-681, 2022.

[14] J.Wang. “Lightweight and real-time object detection model on edge devices with model quantization,” Journal of Physics: Conference Series, vol. 1748, no. 3, pp.1-10, 2021.

[15] B. Jacob, S. Kligys, B. Chen, et al. “Quantization and Training of Neural Networks for Efficient Integer-arithmetic-only Inference” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2704-2713, 2018.

[16] J. Achterhold, J. M. Koehler, A. Schmeink, et al. “Variational Network Quantization,” International Conference on Learning Representations,, pp. 1-18, 2018.

[17] M. Rastegari, V. Ordonez, J. Redmon, et al. “Xnor-net: Imagenet Classification Using Binary Convolutional Neural Networks,” European Conference on Computer Vision, Springer, Cham, pp. 525-542, 2016.

[18] Z. Liu, B. Wu, W. Luo, et al. “Bi-real Net: Enhancing the Performance of 1-bit Cnns with Improved Represent-ational Capability and Advanced Training Algorithm,” Proceedings of the European Conference on Computer Vision(ECCV), pp. 722-737,2018.

[19] Z. Liu, Z. Shen, M. Savvides, et al. “Reactnet:Towards precise binary neural network with generalized activation functions,” European conference on computer vision, Springer, Cham, pp. 143-159, 2020.

[20] E. Soufleri and K. Roy, “Network Compression via Mixed Precision Quantization Using a Multi-Layer Perceptron for the Bit-Width Allocation,” IEEE Access, vol. 9, pp. 135059-135068, 2021.

[21] D. Lin, S. Talathi, S. Annapureddy, “Fixed point quantization of deep convolutional networks,” International conference on machine learning, pp. 2849-2858, 2016.

[22] K. Wang, Z. Liu Z, Y. Lin, et al. “Haq: Hardware-aware Automated Quantization with Mixed Precision,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8612-8620, 2019.

[23] R. Q. Wang, et al. “Deep Neural Network Compression for Plant Disease Recognition,” Symmetry, vol. 13, no. 10, pp.1-17, 2021.

[24] P. J. He, Z. Wu, S. Zhang, et al. “Deep network quantization via error compensation,” IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 9, pp. 4960-4970, 2022.

[25] Z. Dong, Z. Yao, A. Gholami, et al. “Hawq: Hessian Aware Quantization of Neural Networks with Mixed-precision,” Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 293-302 ,2019.

[26] Z. Dong, Z. Yao, D. Arfeen, et al. “Hawq-v2: Hessian Aware Trace-weighted Quantization of Neural Networks,” Advances in Neural Information Processing Systems, vol. 33, pp. 18518-18529, 2020.

[27] Y. Yu, C. Wu, T. Zhao, K. Wang and L. He, “OPU: An FPGA-Based Overlay Processor for Convolutional Neural Networks,” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 28, no. 1, pp. 35-47, 2022.

[28] J. Wang, S. Li, Z. An,et al. “Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines,” *Neurocomputing*, 2018, vol. 329, pp. 53-65.

[29] G. Li, “Block Convolution: Toward Memory-Efficient Inference of Large-Scale CNNs on FPGA,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 5, no. 41, pp.1436-1447, 2022.

[30] M. Cho, Y. Kim. “FPGA-Based Convolutional Neural Network Accelerator with Resource-Optimized Approximate Multiply-Accumulate Unit,” Electronics, vol. 10, no. 22, pp. 1-16, 2021.

[31] M. Sait. “Optimization of FPGA-based CNN accelerators using metaheuristics,” The Journal of Supercomputing, vol. 79, no. 4, pp. 4493-4533, 2023.

[32] P. Tommaso, R. Emilio, D. Gianmarco, et al. “A Multi-Cache System for On-Chip Memory Optimization in FPGA-Based CNN Accelerators,” Electronics, vol. 10, no. 20, pp. 1-18, 2021.

[33] Gao Z, Zhang H, Yao Y, et al. “Soft error tolerant convolutional neural networks on fpgas with ensemble learning,” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 30, no. 3, pp. 291-302, 2022.

[34] J Hu, Ying. G, Q. Tian, et al. “Hardware implementation of neural network accelerator based on RISC-V,” Electronics & Packaging, vol. 23, no. 2, pp. 1-6, 2023.

[35] Abtahi T , Shea C , Kulkarni A , et al. “Accelerating Convolutional Neural Network With FFT on Embedded Hardware,” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, pp. 1-24, 2018.

[36] Hosseini M, Mohsenin T, et, al. “Binary Precision Neural Network Manycore Accelerator,” ACM Journal on Emerging Technologies in Computing Systems(JETC), pp . 1-27, 2021.

[37] Y. Yu, P. Zhang, H. Gong, et al. “Lightweight Network Hardware Acceleration Design for edge computing,” Computer Science, vol. 50, no. S2, pp. 832-838, 2023.

Combining Unsupervised and Supervised Learning to Predict Poverty Households in Sakon Nakhon, Thailand

Sutisa Songleknok, Suthasinee Kuptabut*
Department of Computer, Sakon Nakhon Rajabhat University, Sakon Nakhon, Thailand

***Abstract*—Poverty is a problem that various government agencies are attempting to address accurately and precisely. This solution relies on data and analysis of features affecting poverty. Machine Learning is a technique to analyze and focus on poverty features encompassing five livelihood capitals: human, physical, economic, natural, and social capital to understand the household context and environment. The dataset contains 1,598 poverty households from Kut Bak district, Sakon Nakhon, Thailand. K-prototype was used to group categorical and numerical dataset into four clusters and labelled as Destitute, Extreme poor, Moderate poor, and Vulnerable non-poor. The performances of the Decision tree classifier with feature selection algorithms, including MI, ReliefF, RFE, and SFS, are compared. The best performance is SFS with F-measure, precision, and recall at 74.6%, 74.8%, and 74.7%, respectively. The result is the decision tree rules to predict the poverty level of households, enabling the establishment of guidelines for resolving household issues, and addressing broader problems within the areas.**

Keywords—K-prototype; decision tree; feature selection; Sakon Nakhon poverty households; unsupervised learning; supervised learning

I. INTRODUCTION

Poverty alleviation policy holds a significant role in every nation, with each country tailoring its poverty criteria to evaluate household poverty accordingly. Specifically, many of these countries utilize the poverty line measurement, a criterion established by the United Nations [1], to gauge household poverty levels. Subsequent poverty-related issues arise from various factors, extending beyond just income. Pandemic mitigating investments have largely not rectified issues such as poor health, lack of quality agricultural resources, and production quantity. These factors have further contributed to household poverty, making them multidimensional poor.

Thailand has established an information system dedicated to poor households. Responsible organizations, namely the Community Development Department [2] under the Health Board of Quality of Public Life Development (HBL), oversee data collection, database management, and information display through official online platforms. This organization collects basic minimum needs data at a household level in Thailand. This data demonstrates household members' fundamental status across various aspects of life quality, adhering to minimum standards.

In collaboration with the Office of the National Economics and Social Development Council (NESDC) and the National Electronics and Computer Technology Center (NECTEC), the Thai People Map and Analytics Platform (TPMAP) [3], was developed. This platform is designed to identify individuals in impoverished households who meet the criteria for basic minimum needs across five dimensions of poverty: health, living conditions, education, income, and access to public services. In addition, the Program Management Unit on Area-Based Development (PMUA) has developed a system called Practical Poverty Provincial Connext (PPPConnext) [4] to collect data on impoverished households within 20 provinces, which rank lowest in the nation's Human Achievement Index (HAI) regarding income. This data collection focuses on five dimensions of livelihood assets: human capital, economic capital, natural capital, physical capital, and social capital. These dimensions are utilized to develop appropriate solutions that match the specific needs of these households.

Most studies on poverty primarily focus on the target group at a household level, employing various research methodologies, including qualitative, quantitative, and Machine learning techniques. The poverty data were analyzed using two techniques: 1) Supervised Learning: This approach involves formulating the poverty level or target class. For instance, the research in [5], a study on factors affecting poverty, in [6], the prediction of households exhibiting characteristics at risk of poverty, and in [7] the development of prediction models for depression levels among the elderly in low-income households. This model utilizes techniques, such as Decision tree, Logistic regression, Neural networks, and Random forest. 2) Unsupervised Learning: This approach involves processing data without predetermined poverty level formulations, such as the research in [8, 9] that used clustering techniques to categorize impoverished households. Poverty alleviation programs should be prioritized in household clusters based on the poor conditions identified within each cluster. The research in [10] analyzes the factors affecting the sustainable livelihoods of poverty households. Both techniques contribute to predicting the relationship between factors and poverty, shaping policy guidelines and effective solutions to alleviate poverty.

The objectives of this paper are 1) to cluster the poverty households, 2) to compare the performance of feature selection techniques using a Decision tree classifier, and 3) to create rules to predict poverty status households. We utilized data from impoverished households collected through PPPConnext, and applied data mining techniques to comprehend the characteristics associated with poverty. The identified features are analyzed using both unsupervised learning and supervised learning models. Unsupervised learning, specifically the K-prototype algorithm, is employed to cluster the households based on their living capital aspects. Supervised learning, including Mutual information (MI), ReliefF, Recursive feature elimination (RFE), and Sequential forward selection (SFS) are utilized to reduce the information sizes. A Decision tree is utilized to construct a comparative model, select appropriate features, explain household characteristics, and predict the poverty level of households effectively.

The paper is structured as follows: Section II provides a comprehensive review of the related works; Section III outlines the proposed framework, detailing both clustering and classification steps, and Section IV presents the obtained results. Section V and Section VI wrap up the paper with a discussion and conclusion, respectively.

II. Related Work

A. Multidimensional Poverty

Poverty can be defined from different perspectives. Organizations, such as NESDC use the “poverty line” to establish standards for basic minimum food needs and essential goods, quantified in Baht per person per month. The poverty line has changed accordingly over the years. Individuals earning less than this threshold are classified as “poor.” This categorization is determined by comparing monthly income against the poverty line [11, 12]. Poverty has a relationship with the households’ economic status. The measurement of poverty in general uses the poverty line criteria as the condition for classifying poverty and non-poverty of households, which is widely used at provincial, national, and global levels. Nevertheless, extensive global research revealed that household poverty can be attributed to either insufficient income or various other contributing factors.

The UNDP and the Oxford Poverty and Human Development Initiative (OPHI) [11] have jointly created the Multidimensional Poverty Index (MPI) comprising 10 dimensions: nutrition, child mortality, and years in schooling, school attendance, cooking fuel, sanitation, drinking water, electricity, housing, and asset ownership. A lower MPI value for a country signifies reduced poverty, whereas a high MPI suggests significant multidimensional challenges, related to inequalities in areas like gender, ethnicity, and infrastructure.

B. Feature Selection

Feature selection (FS) involves the process of selecting, removing, and reducing duplication of the relevant features. There are three approaches to feature selection: Filter, Wrapper, and Embedded [13 - 15].

1) Filter approach. Filters evaluate the relevance of features based on intrinsic characteristics. The popular filter approaches include MI and ReliefF.

2) Wrapper approach. The wrapper approach constructs prediction models considering the feature interactions. The well-known wrapper approaches are REF and SFS.

3) Embedded approach. In feature selection, three main approaches are employed: Filter, Wrapper, and Embedded. Embedded approaches like LASSO (L1-Regularization) and RIDGE (L2-Regularization) combine aspects of both filters and wrappers.

In this paper, feature selection was implemented using the Filter approach (MI and ReliefF) and the Wrapper approach (REF and SFS).

C. Unsupervised Learning

The clustering technique, an unsupervised learning approach, is used for categorizing data based on similarities in their characteristics. In the analysis of poverty, this technique segregates impoverished households according to dataset features. The parameter k (number of clusters) can be determined either through predefined business rules, or varied techniques to determine an appropriate value for “k”.

When employing the clustering technique with the numeric dataset, the utilization of K-means clustering is required. For instance, the study in [8] applied the K-means algorithm to assess poverty status and categorize it into three levels: low, medium, and high poverty. This method was implemented in households in Hulu Sungai Tengah Regency, Indonesia. The study’s findings would be utilized to develop policies tailored to individual households to achieve specific goals. The study in [9] analyzes poverty conditions within a community in the Philippines, and groups households into three clusters: stable, critical, and at-risk. Each cluster offers valuable insights into poverty conditions, guiding the community’s planning and program implementation.

In study [16], the utilization of the clustering technique to group poverty households of Lagangilang, Abra, Philippines was divided into three clusters: non-poor, near poor, and poor. Each group describes different characteristics of households following health and nutrition, education, income, and livelihood to formulate appropriate poverty reduction programs. In cases [10] involving both numerical and categorical data, the K-prototype algorithm was employed. For example, this approach was utilized to analyze factors influencing the sustainable livelihoods of impoverished households, using data mining from poor households in Kut Bak district in Sakon Nakhon province, Thailand. The dataset classified poor households into four clusters: extreme, high, moderate, and low poverty levels.

D. Supervised Learning

Classification, a supervised machine learning method, involves constructing models to predict data labels in a given dataset. For example [17], poverty prediction was carried out using three techniques: Softmax classification, Random forest classification, and Multi-layer perception classifier. In predicting impoverished household data from the Cambodia DHS dataset, two types of predictive outcomes are considered: three-class classification (poor, middle, rich), and five-class classification (poorest, poorer, middle, richer, richest). The study revealed that the three-class classification achieved a higher accuracy of 87%, compared to the five-class classification.

Another research in [18] focused on identifying the causes of old-age poverty in South Korea, the decision tree algorithm was applied using 13 variables, with old-aged poverty as the target. The study revealed that earned income was the most significant factor influencing elderly poverty. In another research [19] conducted in Malaysia, Naive Bayes, Decision tree, and K-nearest neighbors' classifiers were employed to predict the bottom 40 percent of poverty households. Among these models, the decision tree model achieved the highest performance. Additionally, in [20] the decision tree model was utilized to predict household poverty based on health status using the Cuatro Santos health and demographic surveillance databases in Nicaragua. The key indicators of poverty, such as the presence of piped water with a meter, the highest education level in households, and ownership of a refrigerator.

This paper employed the Decision tree algorithm to generate a tree-like structure to represent classification rules. In this structure, internal nodes represent dataset features, branches represent decision rules, and each leaf node represents a specific class label.

III. METHODOLOGY

We acknowledge the importance of outlining the reasons behind our selection of the "proposed framework" in addressing the specific problems at hand. Here are the reasons that make the proposed framework appropriate for addressing such problems. Collectively, these facets enable comprehensive analysis, feature emphasis, and effective model evaluation, making the framework apt for addressing the problem. The proposed framework follows a systematic four-phase structure, ensuring methodical handling from data preprocessing to model evaluation. It sources the poverty dataset from a reliable source (PPPConnext) and employs K-prototype clustering for grouping similar data elements, aiding focused analysis. Automatic labelling of poverty households within clusters streamlines interpretation. Multiple feature selection techniques (MI, ReliefF, REF, and SFS) enhance model efficiency by emphasizing impactful attributes. Evaluation via Decision tree classifier comparing feature sets based on F-measure, precision, and recall ensures a robust model selection.

The studies as mentioned earlier used either unsupervised or supervised learning techniques. However, our framework utilized both unsupervised and supervised learning. Unsupervised learning was used to determine the poverty status of households, while supervised learning selected suitable feature datasets to generate rules using a decision tree model for predicting poverty status.

The limitations of the existing framework that may hinder its suitability for the current problems include the dependency on specific datasets restricting adaptability to new data structures, the clustering method may struggle to effectively group elements in different data types, limited feature selection techniques hindering the identification of crucial attributes, rigid evaluation metrics might not align with the problem domain, and inflexibility in model selection might limit adaptability to diverse data demands.

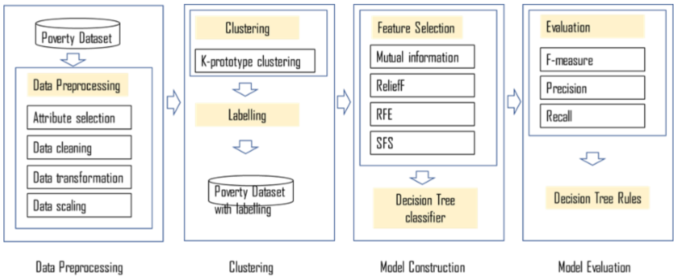


Fig. 1. Conceptual framework.

The conceptual framework shown in Fig. 1 outlines a four-phase process: Data preprocessing, Clustering, Model construction, and Model evaluation. In the data preprocessing phase, the poverty dataset was sourced from the PPPConnext database and underwent preparation for analysis. Utilizing K-prototype clustering, data objects with similar characteristics were grouped into clusters; ensuring data with differing characteristics were placed in distinct clusters. The poverty households cluster was then automatically labelled. Following this, feature selection techniques, MI, ReliefF, REF, and SFS, were applied to identify relevant features. The study then evaluated the performance of the Decision tree classifier by comparing various feature sets, selecting the one with the best performance based on F-measure, precision, and recall.

A. Data Set

The dataset utilized in this study originates from the PPPConnext database, focusing on poor households situated in Kut Bak district, Sakon Nakhon province, Thailand. Sakon Nakhon province ranked 71st out of 78 provinces in terms of income index in 2019. Within this province, Kut Bak district had a poverty rate reaching TPMAP in 2019 at the highest level. This district also held the distinction of having the lowest income level in Sakon Nakhon province. The dataset consists of five types of livelihood assets (human capital, physical capital, economic capital, natural capital, and social capital) to explain the asset limitations of poverty and power within the households. In total, the dataset contains responses from 1,598 households with 76 features (58 categorical features and 18 numerical features). These features are described in detail in Table I.

TABLE I. DETAILED DESCRIPTION OF FEATURES

capitals	**Attributes**	**Explanation**	**Value**
Economic capital (24 features)	income_remitted (N)	monthly remittances	Mean:909.57
	income_farming (N)	monthly income from farming	Mean:3,074.45
	income_non_farming (N)	monthly non-farming income	Mean:6,074.45
	income_welfare (N)	monthly income from state welfare	Mean:893.93
	expenses (N)	monthly household expenses	Mean: 6,951.47
	rice_farming (C)	rice farming households	0=No (14.02%) 1=Yes (85.98%)
	livestock (C)	livestock raising households	0=No (97.43%) 1=Yes (2.57%)

capitals	Attributes	Explanation	Value
	fishing (C)	freshwater fishery households	0=No (79.91%) 1=Yes (20.09%)
	industrial_crop (C)	industrial crop farming households	0=No (86.92%) 1=Yes (13.08%)
	poultry (C)	poultry farming household	0=No (94.62%) 1=Yes (5.38%)
	pig (C)	pig farming household	0=No (99.12%) 1=Yes (0.88%)
	cattle (C)	cattle farming households	0=No (72.03%) 1=Yes (27.97%)
	loan_cousin (C)	a loan from cousins with no collection of interest	0=No (98.56%) 1=Yes (1.44%)
	loan_cousin_ interest (C)	a loan from cousins with a collection of interest	0=No (95.24%) 1=Yes (4.76%)
	loan_community (C)	a loan from the financial savings community-based organizations	0=No (78.85%) 1=Yes (21.15%)
	loan_state (C)	a loan from state-support financial savings	0=No (78.29%) 1=Yes (21.71%)
	loan_BAAC (C)	a loan from the Bank for Agriculture and Agricultural Cooperatives	0=No (82.67%) 1=Yes (17.33%)
	loan_savings_ bank (C)	a loan from the Government Savings Bank	0=No (97.62%) 1=Yes (2.38%)
	loan_commercial _bank (C)	a loan from the Thai Commercial Bank	0=No (98.81%) 1=Yes (1.19%)
	loan_private (C)	a loan from Private AMC	0=No (98.87%) 1=Yes (1.13%)
	loan_creadit_shop (C)	households with credit accessibility from consumer goods shops and production factors	0=No (99.50%) 1=Yes (0.50%)
	loan_informal_ debt (C)	a loan from informal debt	0=No (99.50%) 1=Yes (0.50%)
	student_loan (C)	a loan from a student loan fund	0=No (98.62%) 1=Yes (1.38%)
	savings (C)	households with savings	0=No (45.74%) 1=Yes (54.26%)
Physical capital (31 features)	water_resources (C)	farming using water from water resources, such as rivers, brooks, and ditches	0=No (65.71%) 1=Yes (34.29%)
	reservoirs (C)	farming using water from reservoirs, and village ponds	0=No (93.43%) 1=Yes (6.57%)
	groundwater (C)	farming using water from under groundwater, surface water, artesian aquifer	0=No (85.61%) 1=Yes (14.39%)
	rainwater (C)	farming using rainwater	0=No (40.05%) 1=Yes (59.95%)
	irrigation_canals (C)	farming using irrigation canals	0=No (97.93%) 1=Yes (2.07%)
	ownership_rights (C)	households with issuing legal land rights documents, such as Title Deed	0=No (66.83%) 1=Yes (33.17%)
	ownership_rights _rai (N)	number of arable lands issuing legal land rights documents of ownership (Rai)	
	government_lease (C)	households with land title documents, such as ALRO, S.K.1	0=No (36.55%) 1=Yes (63.45%)
	government_lease _rai (N)	number of arable issuing land title documents (Rai)	
	no_title_arable_ land (C)	households with no land tile documents for arable lands in forest-protected areas or national parks	0=No (90.68%) 1=Yes (9.32%)
	no_title_arable_ land _rai (N)	number of arable lands within forest-protected areas or national parks (Rai)	
	others_rent_free (C)	households farm on others' arable lands rent-free	0=No (84.29%) 1=Yes (15.71%)
	others_rent_ free _rai (N)	number of households farm on others' arable lands rent-free	

capitals	Attributes	Explanation	Value
	others_rent (C)	households farm on others' arable lands with rent payment	0=No (99.50%) 1=Yes (0.50%)
	others_rent_rai (N)	number of arable lands from others with payment (Rai)	
	no_access_water (C)	arable lands with water resources inaccessible for cultivation	0=No (60.76%) 1=Yes (39.24%)
	fertile_soil (C)	number of arable lands with fertile agricultural land	0=No (86.55%) 1=Yes (13.45%)
	risk_area (C)	arable lands in areas with natural disaster risks, such as floods	0=No (98.62%) 1=Yes (1.38%)
	home_ownership (C)	households owning their own homes	1 = Staying with others (0.63%) 2 = Renting a house (0%) 3 = Building houses on others' lands (6.51%) 4 = Owning their houses (92.87%)
	house_condition (C)	house conditions	1 = Need urgent repair (3.32%) 2 = Need remedial action. (38.05%) 3 = No repair needed. (58.64%)
	house__cleanlines s (C)	cleanliness and organizing belongings of households	0= Messy (7.57%) 1= Not messy (92.43%)
	indoor_sewage (C)	households with an indoor sewage system	0=No (16.90%) 1=Yes (83.10%)
	toilet_sanitation (C)	toilets in households with healthy and sanitation conditions	0=No (5.82%) 1=Yes (94.18%)
	waste_separation (C)	households with waste separation	0=No (9.89%) 1=Yes (90.11%)
	electricity_house (C)	households with electricity	0=No (0.75%) 1=Electricity supplied from another house. (1.56%) 2=Yes (97.68%)
	tap_water (C)	households with water supply	0=No (72.09%) 1=Yes (27.91%)
	drinking_water (C)	households buying drinking water	0=No (21.53%) 1=Yes (78.47%)
	mobile_phone (C)	households having mobile phones	0=No (13.33%) 1=Yes (86.67%)
	computer_ ownership (C)	households with computers	0=No (90.30%) 1=Yes (9.70%)
	IT_welfare (C)	households utilizing technology to access state welfare services	0=No (37.23%) 1=Yes (62.77%)
	IT_income (C)	utilizing technology to increase household incomes	0=No (57.81%) 1=Yes (48.19%)
Human capital (18 features)	member_below_ 15 (N)	households with members' ages ranging over 15 years old	Max:7 Min:0 Mean:0.75
	skills_number (N)	number of household members having diverse skills in professions	Max:2 Min:0 Mean:1.16
	education_level (C)	households having individuals with the highest level of education	0 = No schooling education/non-completion of primary level (4.51%) 1 = Elementary school level (37.30%) 2 = Lower secondary school level (22.47%) 3= Secondary school

capitals	Attributes	Explanation	Value
			level or vocational certificate (23.59%) 4 = Diploma or higher vocational degree (3.82%) 5= Bachelor's degree (7.13%) 6 = Higher than a bachelor's degree (1.19%)
	employed_number (N)	working age in the household (15-59 years)	Max:7 Min:0 Mean:1.99
	farming (C)	farming households	0=No (26.47%) 1=Yes (73.53%)
	general_hired (C)	households with general hired occupation	0=No (77.60%) 1=Yes (22.40%)
	agriculture_ employ (C)	households' employment in the agriculture sector	0=No (91.36%) 1=Yes (8.64%)
	self_employed (C)	households with self-employed	0=No (93.55%) 1=Yes (6.45%)
	fishery (C)	households with fishery	0=No (98.87%) 1=Yes (1.13%)
	civil_services (C)	households with civil services	0=No (94.24%) 1=Yes (5.76%)
	contract_employee _in_government_ sector (C)	household members working as government employees	0=No (96.75%) 1=Yes (3.25%)
	private_employee (C)	household members working in private sectors	0=No (86.11%) 1=Yes (13.89%)
	disabled_number (N)	householders with disabled members with self-reliance	Max:2 Min:0 Mean:0.06
	bedridden_number (N)	household members with bedridden old adults and disabled adults with no self-reliance	Max:2 Min:0 Mean:0.02
	chronic_number (N)	household members with chronic illnesses	Max:4 Min:0 Mean:0.22
	healthy_number (N)	households with healthy members	Max:10 Min:0 Mean:2.96
	welfare_card (C)	households' members possessing public welfare card	0=No (10.95%) 1=Yes (89.05%)
	elderly_number (N)	elderly household members aged over 60 years old	Max:3 Min:0 Mean:0.64
Natural capital (2 features)	natural_living (C)	households using natural resources for livelihood, such as mushrooms, firewood, forest plants, edible insects	0=No (16.08%) 1=Yes (83.92%)
	natural_income (C)	households using natural resources to earn income such as honey, herbs, mushrooms	0=No (32.17%) 1=Yes (67.83%)
Social capital	join_community_ group (C)	households joining a community, such as occupational groups, finance groups, social welfare groups	0=No (2.25%) 1=Yes (97.75%)

Data type: N = numerical; C = categorical

B. Data Preprocessing

Data preprocessing is a crucial phase, aimed at simplifying data complexity and enhancing data quality before applying data mining algorithms. This phase encompasses four activities: attribute selection, data cleaning, data transformation, and data scaling.

1) Attribute selection: The dataset contains numerous attributes, some of which are irrelevant. This phase focuses on reducing the dataset size by eliminating irrelevant attributes. For example, features related to social capital that describe community characteristics and non-significant attributes, such as those indicating households without relevant information, were removed.

2) Data cleaning: Data cleaning involves filling in missing values and enhancing the data process of cleaning by filling in missing values. Numeric features are imputed with the average value from the same group, while categorical features are replaced with constant values. For example, if the land size of ownership rights is null, the null values are filled with the average value of the land size of all ownership rights or adjusted welfare allowance for the elderly on age brackets: individuals aged 60-69 received 600 Baht/month, those aged 70-79 received 700 Baht/month, those aged 80-89 received 800 Baht/month, and individuals aged 90 and above received 1,000 Baht/month.

3) Data transformation: Data transformation is used to convert textual information into numerical values for analysis. For example, "farmer" is represented as 1, while "non-farmer" represents 0, the highest education level of all household members, the count of elderly individuals aged 60 and above, as well as the count of working-age individuals between 15 and 59, are calculated.

4) Data scaling: Min-max normalization [21] is a scaling technique in which value rescaled data in a range of 0 to 1 using the formula in (1). The technique is applied to specific numerical attributes, such as the number of households, income, and expenses.

$$X' = (X - X_{min})/(X_{max} - X_{min}) \qquad (1)$$

where, X_{max} and X_{min} are the maximum and the minimum values of a feature, respectively.

C. Clustering

Clustering is the process of grouping data with similar characteristics, where clusters exhibit higher similarity within and differ from other clusters. In this paper, the K-prototype is applied to dealing with both numerical and categorical data. This algorithm combines numerical and categorical data to form clusters. The clusters represent 4 categories of poverty, ranging from the most impoverished to the less poor: Destitute, Extreme poor, Moderate poor, and Vulnerable non-poor, according to the PMUA classification. Poverty household status is labelled on each record and defined as a target attribute for creating a model. The resulting cluster offers valuable insights for community planning and implementation.

D. Model Construction

The feature selection process aims to reduce data dimensionality by removing irrelevant features. The algorithm, MI, ReliefF, REF, and SFS are used to select features from the dataset for classification. Relevant features or predictive features are selected by removing the irrelevant features. After selecting the features, the Decision tree classifier builds the model based on several predictive features.

1) Feature selection techniques: The goal of feature selection is to find the optimal feature subset. By eliminating irrelevant features, the number of features can be reduced, the accuracy of the model can be improved, and the running time can be reduced [22].

a) MI: MI is used as a measure of the relationship between a feature and the target output. The higher the value, the more strongly relevant between a feature and the target, which suggests that the feature is selected. If the score is 0 or very low, then a feature and the target are weakly relevant [23].

b) ReliefF: ReliefF is a filter method of the feature selection algorithm. It finds the weights of features in the case where y is a multiclass categorical variable. According to the correlations between features and targets, different weights are assigned to each feature, and the feature with a slighter weight greater than a certain threshold will be removed [24].

c) RFE: RFE is an algorithm to select features in a training set that are more relevant in predicting the target output and removing weak features. The RFE works by searching for a set of features by starting with all features in the training dataset and selecting the most significant features by finding a high correlation between features and target output. Such recursion works until the number of remaining features reaches the desired number [25].

d) SFS: SFS is an algorithm that selects features from the set of features and evaluates them for a model iterate number between the different sets by reducing and improving the number of features so that the model can find the optimal performance and results. The SFS starts with one feature and adds more iteratively [26].

2) Classification: A Decision tree is a technique to build a predictive model through two phases: the training phase builds a model from a training set with a labelled target output, and the testing phase finds the quality of the trained model from the testing set without labelled target output. The model is like a tree structure. The nodes represent the features, the branches represent the decision rules, and the leaf node represents a poverty household status.

E. Model Evaluation

The performance is evaluated based on the calculation of F-measure, precision, and recall using the confusion matrix. It consists of four elements: true positive (TP), false positive (FP), false negative (FN), and true negative (TN) [27].

TP is a condition when the observations coming from positive classes are predicted to be positive. TN is a condition when observations from negative classes are predicted to be negative. FP is a condition when the actual observation comes from negative classes but is predicted to be positive. FN is a condition when the actual observation comes from a positive but in a positive-negative predicted class.

The performance of the experiments is represented using precision, recall, and F-measure that are evaluated using Eq. (2) to Eq. (4), respectively.

$$Precision = TP/(TP+FP) \tag{2}$$

$$Recall = TP/(TP+FN) \tag{3}$$

$$\text{F-measure} = (2*\text{precision}*\text{recall}) / (\text{precision}+\text{recall}) \quad (4).$$

IV. Results

A. Clustering

The clustering algorithm divided the households into 4 clusters. Table II summarizes the cluster based on the important attributes. Cluster 4 constitutes the largest group containing 31.9%. While, the smallest cluster is Cluster 3, making up 19.1% of the entire cluster. The distinct characteristics of each group (four groups) according to the features to formulate the label were addressed below.

TABLE II. The Cluster of Poverty Households

Cluster	Instance	Percentage	Cluster name
1	397	24.8	Vulnerable non-poor
2	387	24.2	Extreme poor
3	305	19.1	Moderate poor
4	509	31.9	Destitute
Total	**1,598**	**100.0**	

The PMUA labels are assigned post-clustering, and their validity hinges on the methodology employed for the assignment. We used a procedure considering inherent cluster characteristics and, when available, external domain knowledge for labeling. To validate these labels, we cross-referenced them with established poverty classification criteria and assessed alignment with expected poverty attributes. Additionally, we ensured label consistency within clusters by employing quality assessment metrics, aiming to uphold the accuracy and reliability of the assigned PMUA labels.

Group classification using a clustering algorithm resulted in four clusters as shown in Fig. 2.

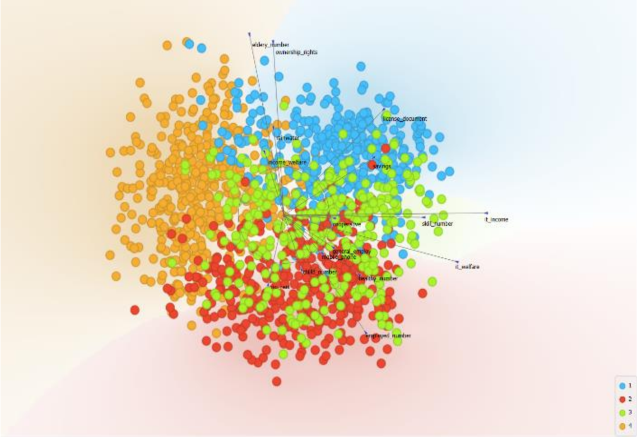


Fig. 2. Freeviz visualization.

Fig. 2 from orange software is used to show the characteristics of four different classes in different colors. Each class describes different characteristics of households. Each axis corresponds to different features; the length of each axis corresponds to the importance of the feature. Cluster 1 is grouped within a blue cluster that is characterized by government lease, savings, and rainwater. Cluster 2 is grouped within a red cluster that is characterized by employed_number, others_rent_free, member_below_15, mobile_phone, general_hired, and healthy_number. Cluster 3 is grouped within a green cluster that is IT_welfare, IT_income, and skills_number, while Cluster 4 is grouped within an orange cluster that is the elderly_number, income_welfare, and ownership_rights. The distribution of instances in Fig. 2 revealed that the instance of Cluster 3 (green area) had fewer colored areas, compared to other clusters. Some instances were mixed up in other clusters and many more were mixed up in Cluster 2 and Cluster 1.

The cluster of the data group divided into four clusters in TABLE II revealed that Cluster 4 had instances more than other clusters. Cluster 1 and 2 had similar cluster stances. Cluster 3 had the least instances. The remaining clusters exhibited similarity in the instance of all clusters, indicating that all householders included members of labor force age, school-age children, elderly, patients with chronic illness, bedridden patients, and disabled members with no self-reliance, disabled members. Most households owned mobile phones. The specific characteristics of each cluster are shown in Fig. 3 as follows:

Cluster 1 included 397 households that were the groups obtaining income from remittances and state welfare programs at the highest mean score, which was more than other clusters. These households also had an average monthly income at the second ranking. The income shows no difference, compared to the cluster with the highest income. The number of elderly and disabled members was at the highest level. The sizes of arable lands with legal land rights documents for ownership, and the legal land rights documents were more than other clusters. Most households completed their education at the elementary school level. The households exhibit the highest savings and use rainwater the most. However, those with features closely aligned to Cluster 3 are referred to as the “Vulnerable non-poor group”.

Cluster 2 included 387 households with an average income from off-farming sectors at the highest level. The average household income from farming sectors and state welfare programs was at the lowest level. The average household expenses ranked second; the labor force age members were at the second-ranking. The arable land sizes were at the lowest rank. Most arable lands had legal rights documents. Most household members completed lower secondary education or vocational certificate education. The households have the highest number of mobile phones, with a quantity similar to Cluster 3. The households have a high number of members under 15 years old, which is also close to that of Cluster 3. This group is referred to as the “Extreme poor group”.

Cluster 3 included 305 households with the average income at the highest level. The average income earned from farming sectors was at the highest level. The income from off-farming sectors was ranked second. Diverse skills in professions were at the highest level. The average number of labor force age and patients with chronic illness were at the highest level. The arable lands of all types were ranked second. Most household members completed their lower secondary education. The households use technology for state welfare applications and income generation, with proximity similar to Cluster 2. This group is referred to as the “Moderate poor group”.

Cluster 4 included 509 households with the average income at the lowest level. The income from off-farming sectors was at the lowest level. The healthy members and the labor force age members were both at the lowest level. The number of elderly was higher than young age members. The diverse skills for professions were at the lowest level. Most household members completed the elementary education level. The average expenses were at the lowest level. The average number of bedridden patients was more than in other clusters but equal to Cluster 1. The average number of chronically ill patients was at the second rank. The average number of disabled members was ranked second to Cluster 1, and the arable lands had the legal land rights documents for ownership. This group is referred to as the “Destitute group”.

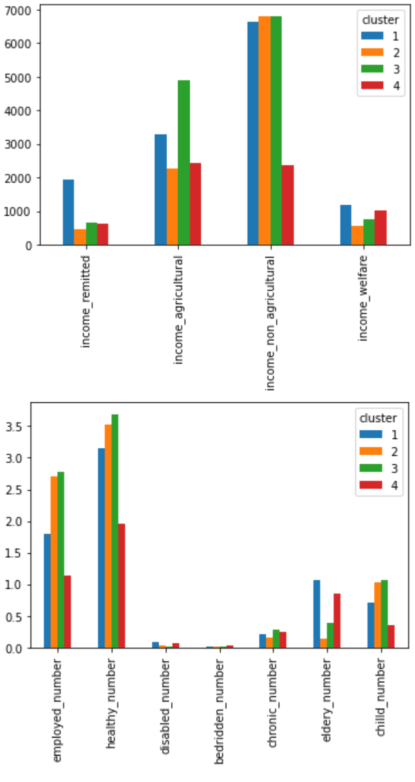


Fig. 3. Data comparison among clusters based on specific characteristics.

B. Feature Selection

The poverty household dataset was classified into four poverty statuses, namely Destitute, Extreme poor, Moderate poor, and Vulnerable non-poor. The values were stored in a column named “target”. Feature selection has been implemented using MI, ReliefF, RFE, and SFS algorithms. These algorithms were implemented in the Scikit-learn library for the selection of optimal features. The number of employed features by the MI algorithm was a high value. The top 20 features were selected to predict the model, shown in Table III. ReliefF calculates a feature score for each feature which can then be applied to rank and select the top 20 scoring features for feature selection, as shown in Table III.

TABLE III. THE SELECTED FEATURES BY EACH ALGORITHM

Features	MI	ReliefF	RFE	SFS	Common
employed_number	✓	✓	✓		✓
IT_welfare	✓	✓	✓	✓	✓
IT_income	✓	✓	✓		✓
ownership_rights_rai	✓	✓			✓
ownership_rights	✓	✓	✓	✓	✓
elderly_number	✓	✓	✓	✓	✓
education_level	✓	✓	✓	✓	✓
skills_number	✓	✓	✓	✓	✓
government_lease_rai	✓	✓	✓		✓
healthy_number	✓	✓	✓		✓
savings	✓	✓	✓	✓	✓
rainwater	✓	✓			✓
no_access_water	✓	✓	✓	✓	✓
expenses	✓	✓	✓		✓
government_lease	✓	✓	✓	✓	✓
cattle	✓	✓	✓		✓
fertile_soil	✓				
income_welfare	✓	✓	✓		✓
loan_state	✓				
rice_farming	✓				
natural_income		✓	✓	✓	✓
home_conditions		✓	✓		✓
child_number		✓	✓		✓
income_farming			✓		
self_employed				✓	
reservoirs				✓	
loan_community				✓	
loan_saving_bank				✓	
loan_commercial_bank				✓	
loan_private				✓	
loan_informal_debt				✓	
student_loan				✓	
welfare_card				✓	
irrigation_canals				✓	
mobile_phone				✓	
bedridden_number				✓	
Total	**20**	**20**	**19**	**21**	**20**

The implementation of the RFE algorithm using the Decision tree classifier on the training set and five cross-validations was performed. The algorithm determined the optimal number of features, and 19 features were selected, indicating the importance of these features on the poverty dataset. The implementation of the SFS algorithm using the Decision tree classifier on the training set and five cross-validations was performed. The algorithm found the best score of 21 features.

Table III shows the selected features that have the most effect on the poverty level. The selected features were considered relevant by at least two selection algorithms. All algorithms selected eight features, and seven features were selected by three algorithms. The five features were selected by 2 algorithms and the other features were selected by only one algorithm. The set of features consists of 20 features, called common features.

C. Model

In this paper, the experiments were divided into 6 experiments that employed different feature selection techniques on a dataset of 1,598 impoverished households.

The selected features gained from the feature selection phase are used to train the Decision tree classifier. The number of features selected by the several types of feature selection algorithms is presented in Table IV. In the processing model, 10-fold cross-validation is applied. Nine folds were used for training and the remaining fold was used for testing. The process was repeated 10 times. The performance metrics such as F-measure, precision, and recall are measured to demonstrate the results and comparative analysis of the feature selection algorithms. The metrics consider the entire features, common features, and the feature set obtained by applying the feature selection techniques. Performance evaluation is given in Table IV.

From Table IV, the performance results varied significantly based on the size of selected features and characteristics attributes. The particular data characteristics associated with SFS using 12 features differed from those of other techniques, thus yielding diverse performance outcomes.

The SFS algorithm has achieved the best performance with F-measure, precision, and recall at 74.6%, 74.8%, and 74.7%, respectively. Thus, the SFS algorithm performed better than other feature techniques. Besides in Table V, a confusion matrix presented the performance of the Decision tree classifier with selected features by SFS algorithm.

TABLE IV. PERFORMANCE OF DECISION TREE CLASSIFIER

Feature selection techniques	No.of features	F-measure	Precision	Recall
All	76	72.0	72.0	72.2
MI	20	71.4	71.5	71.6
ReliefF	20	71.4	71.5	71.6
REF	19	73.6	73.7	73.6
SFS	21	74.6	74.8	74.7
Common features	20	58.9	59.2	58.9

TABLE V. CONFUSION MATRIX OF THE RESULTS OBTAINED BY THE DECISION TREE CLASSIFIER WITH SFS ALGORITHM

Actual	***Predicted***			
	Vulnerable non-poor	***Extreme poor***	***Moderate poor***	***Destitute***
Vulnerable non-poor	74.2%	5.2%	7.8%	6.6%
Extreme poor	6.7%	73.1%	9.8%	7.5%
Moderate poor	8.5%	8.5%	69.5%	6.2%
Destitute	10.6%	13.2%	12.9%	79.7%

From the decision tree model, the total number of trees is 587 nodes, which interprets 293 rules, and the depth is 16. The example tree model showed depths of trees, represented in Fig. 4. The most important feature, is whether using technology to request state welfare benefits is or not (IT_welfare), classified into nodes for with (IT_welfare=1) and without (IT_welfare=0) using technology. The left node shows the high education level of members in the household (education_level), the next level is arable lands with the legal land rights documents (ownership_rights), and the number of older people (elderly_number). The model from the Decision tree classifier can be interpreted and understood in the decision tree rules format. Table VI shows the example of decision tree rules. The relationship rules were able to describe the rule and class.

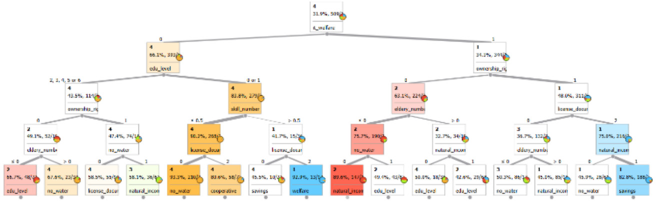


Fig. 4. Decision tree model with 5-level depth.

TABLE VI. DECISION TREE RULES MODEL APPLYING RESULTS

Rules	**Condition (If)**	**Class**	**Description**
1	IT_welfare = 1 and ownership_rights = 1 and government_lease = 0 and elderly_number > 0 and natural_income = 1 and skills_number > 0 and no_access_water = 1 and loan_savings_bank = 0 and education_level in (1,3,4,6) and reservoir = 0	Vulnerable non-poor	IF households utilize technology for requesting the state welfare programs AND have aroma lands with legal land rights documents of ownership AND have aroma lands without legal land rights documents AND have elderly members AND utilize natural resources to earn income AND have skills for professions AND have arable lands with no access to water supply AND no loan from the Savings Banks AND have their highest level of education lower than bachelor's degree AND have access to water supply from reservoirs for farming
2	IT_welfare = 1 and ownership_rights = 0 and elderly_number <= 0 and no_access_water = 0 and natural_income = 1 and education_level in (0,3,4,5,6)	Extremely poor	IF households utilize technology for requesting state welfare programs AND have aroma lands with legal land rights documents of ownership AND no elderly members AND no access to water supply for farming AND utilize natural resources to earn income AND have their highest level of education lower than bachelor's degree
3	IT_welfare = 0 and education_level in (2,3,4,5,6) and ownership_rights = 1 and no_access_water = 1 and natural_income = 0 and savings = 1 and elderly_number <= 0.33	Moderate poor	IF households utilize no technology for requesting the state welfare programs AND have their highest level of education lower than a bachelor's degree AND have aroma lands with legal land rights documents of ownership AND no access to water supply for farming AND no utilizing natural resources to earn income AND have savings AND have at least one elderly member
4	IT_welfare = 0 and ownership_rights = 1 and no_access_water = 0 and government_lease = 0 and savings = 0 and self_employed =0 and education_level = 2	Destitute	IF households utilize no technology for requesting the state welfare programs AND have aroma lands with legal land rights documents of ownership AND no access to water supply for farming AND having no aroma lands with legal land rights documents AND having no savings AND no self-employed business AND have their highest level of education at lower secondary education

V. DISCUSSION

Upon analyzing the factors affecting the classification of poverty levels based on livelihood capital, aiming to construct poverty indicators for classifying poverty levels, it was found that factors derived from the SFS technique and Decision tree classifier yielded the highest F-measure. These factors comprised five factors within human capital, seven factors within physical capital, six factors within economic capital, and 1 factor within natural capital. Nevertheless, there were no factors identified within the social capital category.

Features affecting the predicting of the poverty level, such as education_level, and skills_number, which were features related to the competencies of household members. The study in [28] suggests that the factors influencing poverty in Thailand are linked to the rise in per capita income and education, both showing statistical significance at the 0.10 level. The household members who completed at a basic education level or higher than the basic education level had abilities to earn more income and were able to scaffold knowledge to increasingly improve skills. These could reduce poverty in households.

The feature called IT_welfare for receiving welfare resources from government sectors of Thailand. Some welfare allowances were able to be accessed through the online register, receiving state welfare resources through applications, household members having technological skills and mobile phones. Thus, the household members were able to access state welfare resources and received assistance from the government sectors following the established conditions. The study referenced in [29] shows that the Internet has a significant impact on alleviating the vulnerability to poverty among rural households.

The features, ownership_rights, and government_lease, were related to arable lands of poor households because the land is a critical asset, the primary for generating a livelihood, and a main vehicle for investing, especially for the poor as it provides a means of livelihood through the production and sale of crops and other products. The studies in [30, 31] suggest that land can serve as collateral for credit to invest in the land or be exchanged for capital to start up another income-generating activity. The study in [32] suggests that the absence of land ownership may contribute to a high fertility rate, low capital investment, and consequently lower living standards.

The feature called no_access_water was related to water supply for farming was limited. Most farmers use rainwater as a major waste resource for agriculture. Some areas have water resources close to their arable lands, such as canals, and ponds. Thus, the increasing amount of water for farming areas is important for farmers to be able to farm yearly.

The elderly, children aged from 0-14 years of age, chronically ill, and bedridden patients living in households are vulnerable groups, who are key features affecting poverty because many households, while not currently in poverty, recognize that they are vulnerable to events - a bad harvest, a lost job, an illness, and unexpected expenses, an economic downturn - that could easily push them into poverty [33]. A vulnerability group is a group with no income or few incomes that rely on family members or households with members in the labor force, but rely on the state welfare system, such as the subsidy support budget program for newborn babies to six years old, the subsistence allowance provision for disabled persons. These support household expenses and alleviate poverty for vulnerable groups [34] or households would afford health care if the households had members with inpatients, households with members aged over 65 years, and households with disabilities [35].

VI. Conclusion

The conclusion concisely outlined our key findings, insights, and results, demonstrating the importance and implications of our work in addressing the issue.

The research employed both supervised and unsupervised learning techniques to analyze data on poor households. Initially, unsupervised learning divided the data into four clusters representing different poverty levels: Destitute, Extreme poor, Moderate poor, and Vulnerable non-poor groups. Using supervised learning, the study identified key features influencing poverty levels, employing algorithms like MI, ReliefF, RFE, and SFS. The decision tree model using the SFS algorithm proved most effective, achieving 74.6% F-measure, 74.8% precision, and 74.7% recall in predicting poverty levels. The study then summarized the characteristics of each poverty level in the Kut Bak district:

Destitute Households: These households earn an average monthly income of 6,426 Baht; heavily rely on state welfare due to limited development opportunities, with elderly members, minimal education levels, poor health, and reliance on government services.

Extreme Poor Households: With monthly income averaging around 10,088 Baht, these households mostly derive earnings from non-agricultural sources due to limited arable lands, facing excessive expenses with minimal savings, but possessing skills to access welfare resources.

Moderate Poor Households: These households earn an average income of 13,100 Baht monthly from both farming and off-farming sectors, facing substantial expenses on healthcare, but managing some savings. Lack of water access impacts household development.

Vulnerable Non-Poor Households: This group earns an average of 13,030 Baht monthly, primarily from agricultural sources and remittances, owning considerable arable lands, providing opportunities for an improved quality of life compared to other clusters.

In future work, we will create applications to predict households of poverty in four levels with the conditions drawn from the decision tree and using the class of poor households in the area of Kut Bak district. The leader of the community was a person who selected households following the characteristics of four clusters. After entering the data through the applications for predicting poor households to determine the accuracy between humans and machines and create the acceptance of the results from the predicting with the relevant organization in the areas to apply it for actual practices.

Acknowledgment

This research is part of the “Precision Poverty Alleviation in Sakon Nakhon Province” awarded as a research scholarship from the Program Management Unit on Area Based Development (PMUA) in 2022. Thank you to all households in Kut Bak district, Sakon Nakhon for collecting data.

Ethical Consideration

The data was collected from households in Kut Bak district in Sakon Nakhon located in the northeastern region of Thailand in 2022. Ethics approval for the fieldwork was obtained through the Human Research Ethics Committee at Sakon Nakhon Rajabhat University. Guarantee No. HE 65-095.

References

[1] United Nations, The Sustainable Development Goals Report 2022, New York: United Nations, 2022.

[2] Community Development Department, “Program for Collected, Storage, and Process in 2023 - 2027,” 2023. https://smartbmn.cdd.go.th (accessed Jun. 15, 2023).

[3] National Science and Technology Development Agency, “TPMAP Thai People Map and Analytics Platform,” 2018. https://www.tpmap.in.th/ (accessed Jun. 15, 2023).

[4] Program Management Unit on Area Based Development, “Practical Poverty Provincial Connext: PPPConnext,” 2019. http://www.ppaos.com/ppaos47/frontend/web/ (accessed Aug. 1, 2023).

[5] B. Reddy, K. Srikanya, M. Varshini, S. Srijanya, G. Reddy and C. Shirisha, “The Application of Machine Learning to the Task of Poverty Classification,” 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-4, doi: 10.1109/ACCAI58221.2023.10201169.

[6] H. A. Silva Marchan, O. J. M. Peña Cáceres, D. M. Ricalde Moran, T. Samaniego-Cobo, and C. M. Perez-Espinoza, “A Machine Learning Study About the Vulnerability Level of Poverty in Perú,” Communications in Computer and Information Science, Springer, Cham, vol. 1658, 2022, doi: https://doi.org/10.1007/978-3-031-19961-5_1.

[7] K. M. Kim, J. H. Kim, H. S. Rhee, and B. Y. Youn, “Development of a Prediction Model for the Depression Level of the Elderly in Low-income Households: using Decision Trees, Logistic Regression, Neural Networks, and Random Forest,” Sci Rep, vol. 13, July 2023, doi: https://doi.org/10.1038/s41598-023-38742-1.

[8] N. Istiqamah, O. Soesanto, and D. Anggraini, “Application of the K-Means Algorithm to Determine Poverty Status in Hulu Sungai Tengah,” J. Phys, vol. 2106, 2021, doi: 10.1088/1742-6596/2106/1/012027.

[9] P. R. Michelle and C. R. Rex Aurelius, “Apply Clustering Algorithm on Poverty Analysis in a Community in the Phillippines,” Proceeding of the international conference in industrial engineering and operations management Monterrey, Mexico, 3-5 November 2021, pp. 1511-1521.

[10] S. Kuptabut and S. Songleknok, “The Analysis of Factors Affecting the Sustainable Livelihoods of Poverty Households using Data Mining Technique,” 2022 International Conference on Digital Government Technology and Innovation (DGTi-CON), Bangkok, Thailand, 2022, pp. 20-23, doi: 10.1109/DGTi-CON53875.2022.9849207.

[11] United Nations Development Programme and Oxford Poverty and Human Development Initiative, 2023 Global Multidimensional Poverty Index (MPI): Unstacking global poverty: Data for high impact action, New York: UNDP, 2023.

[12] World Bank, “Poverty & Equity Brief Thailand East Asia & Pacific April 2023,” 2023. https://databankfiles.worldbank.org/ (accessed Jun. 20, 2023).

[13] Y. Bouchlaghem, Y. Akhiat, and S. Amjad, “Feature Selection: a review and comparative study,” E3S Web of Conferences, vol. 351, 2022, doi: https://doi.org/10.1051/e3sconf/202235101046.

[14] A. Alsharkawi, M. Al-Fetyani, M. Dawas, H. Saadeh, and M. Alyaman, “Improved Poverty Tracking and Targeting in Jordan Using Feature Selection and Machine Learning,” IEEE Access, vol. 10, pp. 86483-86497, 2022, doi: 10.1109/ACCESS.2022.3198951.

[15] P. Kambuya, “Better Model Selection for Poverty Targeting through Machine Learning: A Case Study in Thailand,” Thailand and the World Economy, vol. 38, pp. 91-116, 2020.

[16] J. A. Talingdan, “Data Mining using Clustering Algorithm as Tool for Poverty Analysis,” Proceedings of the 2019 8th International Conference on Software and Computer Applications (ICSCA), 2019, pp. 56-59.

[17] G. Wong, “Poverty Prediction and the Identification of Discriminative Features on Household Data from Cambodia,” TechRxiv, pp. 1511-1521, doi: https://doi.org/10.36227/techrxiv.21266226.v1.

[18] L. Soochang and K. Daechan, “Decision Tree Analysis for Prediction Model of Poverty of the Older Population in South Korea,” International Journal of Advanced Culture Technology, vol. 10, no. 2, pp. 28-33, 2022.

[19] N. S. Sani, M. A. Rahman, A. A. Bakar, S. Sahran, and H. M. Sarim., “Machine Learning Approach for Bottom 40 Percent Households (B40) Poverty Classification,” International Journal Advanced Science Engineering Information Technology, vol. 8, no. 4-2, pp. 1698-1705, 2018.

[20] C. Kallestal, E. Blandon Zelaya, R. Pena, W. Perez, M. Contreras, L. A. Persson, O. Sysoev, and K. Ekholm Selling, “Predicting poverty Data mining approaches to the health and demographic surveillance system in Cuatro Santos, Nicaragua,”. Int J.Equity health, vol. 18, no. 165, 2019.

[21] J. Han, M. Kamber, and J. Pei, “Data transformation and data discretization. In Data Mining-Concepts and Techniques; Kaufmann,” M., Ed.; Elsevier: Amsterdam, The Netherlands, 2011, pp. 111–112.

[22] K. Kira and L. A. Rendell, “A practical approach to feature selection,” 9th International Conference on machine learning, 1992, pp 249-256.

[23] J. R., Vergara and P. A. Estévez, “A review of feature selection methods based on mutual information,” Neural Comput & Applic, vol. 24, pp. 175–186, 2014.

[24] M. Robnik-Sikonja and I. Kononenko, “Theoretical and empirical analysis of ReliefF and RReliefF,” Machine Learning, vol. 53, no. 1–2, pp. 23–69, 2003.

[25] E. M. Senan, M. H. Al-Adhaileh, F. W. Alsaade, T. H. H. Aldhyani, A. A. Alqarni, N. Alsharif, M. I. Uddin, A. H. Alahmadi, M. E. Jadhav, and M. Y. Alzahrani, “Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques,” Journal of Healthcare Engineering, vol. 2021, June 2021, doi: https://doi.org/10.1155/2021/1004767.

[26] R. Gutierrez-Osuna, “Pattern Analysis for Machine Olfaction: a review,” IEEE Sensors Journal, vol. 2, no. 3, pp. 189-202, June 2002, doi: 10.1109/JSEN.2002.800688.

[27] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, “Selecting Critical Features for Data Classification based on Machine Learning Methods,” J Big Data, vol. 7, no. 52, July 2020, doi: https://doi.org/10.1186/s40537-020-00327-4.

[28] S. Chotikhamjorn, “The Study of Poverty and Factors Affecting Poverty in Thailand,” Management Science Review, vol. 23, pp. 63-71, July – December 2021.

[29] S. Zhang, Q. Liu, X. Zheng, and J. Sun, “Internet Use and the Poverty Vulnerability of Rural Households: From the Perspective of Risk Response,” Sustainability, vol. 15, no. 2, 2023.

[30] R. Meinzen-Dick, DESA Working Paper No. 91: Property Rights for Poverty Reduction?, New York: United Nations, 2009.

[31] K. Deininger, Land and Policies for Growth and Poverty Reduction: a World Bank policy research report, Washington: Oxford University, 2003.

[32] L. Cellarier, “Is land ownership a ladder out of poverty?” World Development, vol. 146, 2021.

[33] P. Lant, S. Asep, and S. Sudarno, Quantifying Vulnerability to Poverty: A Proposed Measure, Applied to Indonesia. Policy Research Working Paper; No. 2437, World Bank: Washington, DC. 2000.

[34] T. Jiratitikulchai, S. Lilehhannot, and P. Fungpiriya, Discussion Paper No.67: Economic Status of Vulnerable Households and Considerations for Social Protection. Bangkok: Thammasat university, 2022.

[35] N. Wang, J. Xu, M. Ma, L. Shan, M. Jiao, Q. Xia, W. Tian, X. Zhang, L. Liu, Y. Hao, L. Gao, Q. Wu, and Y. Li, “Targeting Vulnerable Groups of Health Poverty Alleviation in Rural China— What is the Role of the New Rural Cooperative Medical Scheme for the Middle Age and Elderly Population?,” Int J Equity Health, vol. 19, pp. 1-13, 2020.

Network Oral English Teaching System Based on Speech Recognition Technology and Deep Neural Network

Na He[1*], Weihua Liu[2]
School of Foreign Languages, Pingxiang University, Pingxiang, 337000, China[1]
Pingxiang Branch of Jiangxi Telecom Company, Pingxiang, 337000, China[2]

***Abstract*—With the development of computer technology, computer-aided instruction is being used more and more widely in the field of education. Based on speech recognition technology and deep neural network, this paper proposes an online oral English teaching system. Firstly, the speech recognition technology is introduced and its feature extraction is elaborated in detail. Then, three basic problems and three basic algorithms that need to be solved in speech recognition system using Markov model are discussed. The application of HMM technology in speech recognition system is studied, and some algorithms are optimized. The logarithmic processing of Viterbi algorithm, compared with the traditional algorithm, greatly reduces the amount of computation and solves the overflow problem in the operation process. By combining deep network with HMM, continuous speech signal modeling is realized. According to the characteristics of the DNN-HMM model, it is proposed that the model cannot model the long-term dependence of speech signals and train complex problems. Based on Kaldi, the model training comparison experiments of monophonon model, triphonon model and adding feature transformation technology are carried out to continuously improve the model performance. Finally, through simulation experiments, it is found that the recognition rate of the optimized DNN-HMM mixed model proposed in this paper is the highest, reaching 97.5%, followed by the HMM model, which is 95.4%, and the lowest recognition rate is the PNN model, which is 90.1%.**

Keywords—Deep neural network; Markov model; voice design technology; Viterbi algorithm; oral English teaching

I. INTRODUCTION

With the evolution of computer science and technology, computer-aided instruction is being used more and more widely in the field of education. Nowadays, with the help of computer-aided instruction, people can learn languages more conveniently [1]. The rich graphics and sound processing functions of the computer effectively promote the language learning effect of people. At present, research hotspots in this field focus on exploring effective language learning methods that combine speech recognition technology with multimedia technology [2]. The development of software for teaching spoken English with speech identification has emerged as a hot topic in this type of language teaching.

For the time being, given the cutting-edge lookup development of monosyllabic recognition, Zhang Jing et al. first added the algorithm primarily based on finite country vector quantization and the lookup consequences of its expanded algorithm in monosyllabic recognition, then added the algorithm based totally on the implicit Markov model, and special brought the lookup consequences of syllable attention combining hidden Markov mannequin with different administration strategies [3]. Yang et al. introduced the Fisher criterion and L2 regularization constraint to ensure the minimization of parameter errors in the stage of backpropagation adjustment of parameters, the dispersion of samples between classes and the concentration of intra-class distributions after classification, and the proper order of magnitude of network weights to effectively alleviate the overfitting problem [4]. Sun et al. utilized the end-to-end technological know-how primarily based on hyperlink timing classification to Japanese speech recognition. Considering the traits of Hiragana, katakana, and kanji in more than one writing type in Japanese, they explored the effect of special modeling gadgets on awareness overall performance via experiments on Japanese records units [5]. Huang et al. proposed a finite local weight shared convolutional neural network (CNN) speech recognition based on the Meir spectral coefficient (MFSC) feature to address the problem of unsatisfactory recognition effect in traditional speech recognition applications [6]. Hou Yimin et al. mainly analyzed and summarized several current representative deep learning models, introduced their applications in speech identification for speech feature extraction and acoustic modeling, and finally summarized the problems faced before and the development direction [7].

Compared with the traditional single model for speech classification, this paper innovatively proposes to optimize the deep neural network and fuse it with the Markov model, and the recognition rate of the DNN-HMM fusion model is greatly improved. The application of HMM technology in speech recognition system is studied, and some algorithms are optimized. The logarithmic processing of Viterbi algorithm, compared with the traditional algorithm, greatly reduces the amount of computation and solves the overflow problem in the operation process. By combining deep network with HMM, continuous speech signal modeling is realized. According to the characteristics of the DNN-HMM model, it is proposed that the model cannot model the long-term dependence of speech signals and train complex problems.

*Corresponding Author.

II. Speech Recognition Technology

A. Speech Recognition Technology and Feature Extraction

Speech attention means the ability to transform voice symbols into corresponding textual content [8]. Fig. 1 shows the shape of conventional voice attention and it consists in most cases of five parts: feature extraction, acoustic modeling, pronunciation lexicon, language modeling, and decoding search.

The mathematical description of this process is shown in Eq. (1):

$$\hat{W} = \arg\max_{W} P(W|\ O) \tag{1}$$

W is the candidate word sequence.

From Bayes' formula, Eq. (1) could be further written as:

$$\hat{W} = \arg\max_{W} P(W|\ O) = \arg\max_{W} \frac{P(W)P(O|\ W)}{P(O)} \tag{2}$$

P(W) stands for language model and represents the probability of occurrence of word sequence W; P(O|W) stands for acoustic model, which represents the probability of generating feature sequence O given word sequence W; P(O) represents the likelihood of watching the acoustic characteristic O, whose price has no impact on the closing cognizance result and could be overlooked [9]. So, system in Eq. (2) can be written as below:

$$\hat{W} = \arg\max_{W} P(W)P(O|\ W) \tag{3}$$

The speech signal is a kind of non-stationary signal, so it cannot use the traditional signal processing method. It is found that the characteristics of speech signals remain relatively stable in a short period (10~30ms). Therefore, it is necessary to do some processing before feature extraction, that is, pre-emphasis, frame, and window.

The speech signal is attenuated by 12dB/ octave after it is emitted from the glottis and by 6dB/ octave after it is radiated through the mouth. Therefore, in the speech spectrum generated after short-time FFT, the components of the high-frequency part are smaller, and the entire spectrum becomes steeper [10]. To flatten the signal spectrum, a pre-weighting process is generally required to improve the high-frequency portion at a rate of 6 dB per octave. This is usually accomplished using a first-order high-pass numerical filter:

$$H(z) = 1 - \mu z^{-1} \tag{4}$$

Frame processing is primarily used to enable the extraction of acoustic features and is predicated on the short-term stability of speech signals. The technique of overlapping segmentation is used to carry out frame segmentation, allowing for a seamless transition between different voice frames to guarantee their temporal continuity [11]. In the realm of digital signal processing, the most often utilized window functions are the rectangular window and the Hamming window, among others. In particular, the most popular window for voice recognition is the Hamming window, which may significantly reduce the spectrum leakage brought on by the truncation effect [12]. It has the following window functions:

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\dfrac{2\pi n}{N-1}\right) & 0 \le n \le N-1 \\ 0 & \text{others} \end{cases} \tag{5}$$

The human ear perceives different frequency components of speech signals to different degrees, in which it is more sensitive to the low-frequency part and less distinguishable from the high-frequency part. MFCC is designed primarily based on this auditory grasp attribute of the human ear. Fig. 1 suggests the ordinary method of extracting MFCC features:

1) The enter voice sign is preprocessed to attain the time-domain sign after including the window.

2) Due to the challenging nature of analyzing the features of speech symbols in the temporal region, they are usually transformed into the spectral domain for evaluation [13]. The linear spectrum of the time-domain symbols is obtained by a short-time FFT transform:

$$S(k) = \sum_{n=0}^{N-1} s_w(n) e^{-j\frac{2\pi nk}{N}} \quad (0 < k < N) \tag{6}$$

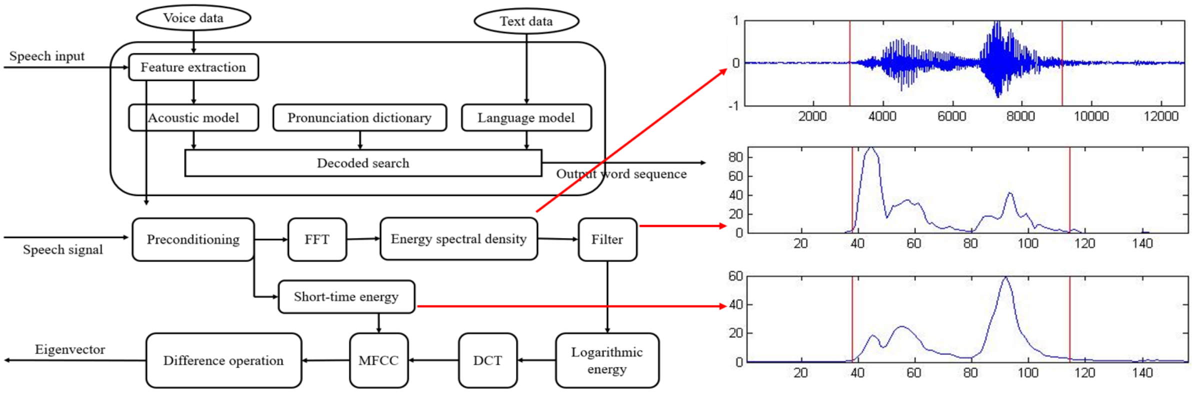


Fig. 1. Speech recognition architecture diagram.

After FFT transformation, the short-time emission profile can be obtained directly:

$$P(k) = S(k)S^*(k) = | S(k) |^2 \tag{7}$$

3) The MFCC features are constructed based on the auditory perception of the human ear, and the Mel frequency corresponds to the perceived frequency of the human ear. Therefore, to convert the linearity spectrum to Mel frequency, it is mainly realized by a set of delta strip bandpass filters uniformly distributed on the Mel frequency ruler [14].

The expression for this set of delta band pass factors is shown in Eq. (8):

$$H_m(k) = \begin{cases} \frac{k - f(m-1)}{f(m) - f(m-1)}, f(m-1) \le k \le f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)}, f(m) \le k \le f(m+1) \\ 0, \text{ others} \end{cases} \tag{8}$$

The conversion formula from linear frequency to Mel frequency is as follows:

$$Mel(f) = 1125 \times \ln\left(1 + \frac{f}{700}\right) \tag{9}$$

Or:

$$Mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \tag{10}$$

In this case, the triangle bandpass filter bank serves the following purposes: it may minimize the quantity of feature data and make calculations easier; it can smooth the spectrum and remove the effects of harmonics.

4) Determine the subband energies at the Mel scaling, or the logarithmic quantities of the energies of each factor in the channel bank, and perform a discharge cosine shift on them.

$$E_m = \ln\left(\sum_{k=0}^{N-1} P(k) H_m(k)\right) \tag{11}$$

$$C_d = \sum_{m=0}^{M-1} E_m \cos\left[m\left(k - \frac{1}{2}\right)\frac{\pi}{M}\right], d = 0,1,\ldots,D-1 < M \tag{12}$$

After a series of operation steps, a frame of speech signal can be represented by a multi-dimensional MFCC vector. For Fbank features, its extraction method is very similar to that of MFCC, except that DCT transformation is not required, so the correlation information between various dimensions is completely retained [15].

Short-time average zero crossing rate applications that can play the following roles in endpoint detection applications:

Distinguish between voiceless and voiceless sounds by the cost of the zero-crossing rate. The corresponding speech section with a greater zero crossing charge is voiceless, whilst the corresponding speech phase with a decreased zero crossing price is dulled [16]; Judging the beginning of speech by the price of the zero-crossing rate, the usage of the zero-crossing charge can be aware of that that is the noise, that is the actual beginning factor of speech; The zero-crossing rate combined with short-time energy can also be used as a basis for judging whether there is speech generation.

B. *Linking Temporal Classification Algorithm*

Both GMM-HMM and DNN-HMM belong to the mixed model in which a variety of models work together. There are many problems in this mixed modeling method. For example, in terms of model training, the process of fully training a hybrid modeling system is very complicated, and the relatively independent training of each module will make it difficult for the system to carry out overall optimization. Moreover, before training the DNN-HMM system, it is necessary to train a GMM-HMM to obtain the frame-level correspondence between speech data and labels, which leads to the final recognition accuracy of the system to a certain extent depending on whether the GMM-HMM model alignment is accurate [17]. In terms of the structural characteristics of the model itself, to obtain a high recognition rate in continuous speech recognition tasks with a large vocabulary, only the modeling units below the word can be selected, such as mono phonemes, tri phonemes, etc., which need to be escaped by pronunciation dictionaries. In addition, the HMM model used to model speech sequence information fails to take into account the contextual correlation between speech frames.

Proposed by Alex Graves in 2006, the linked time sequence classification method is a key technology for end-to-end speech recognition. Neural networks alone complete the entire process of continuous speech recognition, which belongs to an integrated modeling method [18]. The final text sequence can be generated directly, which solves the problems of traditional methods such as forced alignment, cumbersome recognition process, and non-consistent optimization.

Assuming that the conditions between the output symbols of each speech frame are independent, for a single sample (O, W), the final optimization goal of CTC can be described as minimizing the negative logarithm of the posterior probability of the output symbol sequence:

$$L_{CTC} = -\ln P(\mathbf{W} | \mathrm{O}) = -\ln \sum_{\pi \in B^{-1}(\mathbf{W})} \prod_{t=1}^{T} p\left(\pi_t | \mathbf{o}_t\right) \tag{13}$$

From Eq. (13), it can be seen that if the CTC loss function is computed directly, it is necessary to obtain a summation of the conditional probabilities of all possible symbol orders capable of yielding the output character sequence W. Therefore, similar to the forward-backward algorithm in HMM calculation, dynamic programming is usually used to calculate the loss of CTC in practice. By combining the symbol sequence that can get the same output character sequence in the same time step, the calculation amount is reduced and the double calculation is avoided.

$$\alpha(s,t)=\sum_{\pi_{1.t}\in B^{-1}(\mathbf{W}_{1s/2}),\pi_t=w_s'}\prod_{i}^{t}P\left(\pi_i\middle|\ x_i\right) \tag{14}$$

$$\alpha(1,1)=P\left(\varphi\middle|\ o_1\right) \tag{15}$$

$$\alpha(1,2)=P\left(w_1\middle|\ o_1\right) \tag{16}$$

$$\alpha(1,s)=0,\quad \forall s>2 \tag{17}$$

According to the introduction of the neural network, after establishing the neural network, it is necessary to reduce the mistake between the forecasting result of the neural network and the actual sample labeling as much as possible, i.e., to reduce the loss function as much as possible. By analyzing the damage factor of CTC and the calculation method of the loss function, it can be found that the output probability of each time step is microscopic. Therefore, the backpropagation algorithm can be used to update each parameter value of the neural network using the CTC method and minimize the loss function to realize the classification of time series. The following describes the reverse derivation of the CTC loss function on the output of the neural network.

Assume that the number of neurons in the output level of the neural network is |U '|, SoftMax is used as the activation function, and each neuron outputs a posterior probability of a specific symbol on the time step. Then, for a single sample, the bias of the posterior probability of the output character label sequence as a whole to the output of a single neuron of the neural network is:

$$\frac{\partial P(\mathbf{W}|\ \mathbf{O})}{\partial P\left(q\middle|\ o_t\right)}=-\frac{1}{P\left(q\middle|\ o_t\right)^2}\sum_{s\in lab(\mathbf{W},q)}\alpha(s,t)\beta(s,t) \tag{18}$$

Then, the derivative of the CTC loss function to the output of the neural network is:

$$\frac{\partial L_{CTC}}{\partial P\left(q\middle|\ o_t\right)}=\frac{1}{P(\mathbf{W}|\ \mathbf{O})P\left(q\middle|\ o_t\right)^2}\sum_{s\in lab(\mathbf{W},q)}\alpha(s,t)\beta(s,t) \tag{19}$$

III. Markov Model

A. Basic Problem and its Structure

Given the remark sequence and HMM model, if the kingdom transition sequence is known, the chance that the HMM mannequin produces the output remark sequence is:

$$P(O|\ q)=b_{q_1}\left(O_1\right)b_{q_2}\left(O_2\right)\ldots b_{q_T}\left(O_T\right) \tag{20}$$

The probable gas output sequence q of the HMM model is:

$$P(q|\ \lambda)=\pi_{q_1}a_{q_1q_2}a_{q_2q_3}\ldots a_{q_{T-1}q_T} \tag{21}$$

For all possible state transition sequences q, the model outputs the probability of observing sequence O:

$$\begin{aligned}P(O|\ \lambda)&=\sum_{\forall\ q}P(O|\ q,\lambda)P(q|\ \lambda)\\&=\sum_{q_1,q_2,\ldots,q_T}\pi_{q_1}b_{q_1}\left(O_1\right)a_{q_1q_2}b_{q_2}\left(O_2\right)\ldots a_{q_{T-1}-q_T}b_{q_T}\left(O_T\right)\end{aligned} \tag{22}$$

In practice, this amount of computation cannot be borne, so forward algorithm and backward algorithm are adopted to reduce the amount of computation [19].

Define the forward probability as:

$$\alpha_t(i)=P\left(O_1O_2\ldots O_T,q_t=S_i\middle|\ \lambda\right) \tag{23}$$

The forward probability may be computed using the recursion formula that follows:

Starting Point:

$$\alpha_1(i)=\pi_ib_i\left(O_1\right)\quad 1\le i\le N \tag{24}$$

Iteration:

$$\alpha_{t+1}(j)=\left[\sum_{i=1}^{N}\alpha_t(i)a_{ij}\right]b_j\left(O_{t+1}\right),\quad 1\le t\le T-1,1\le j\le N \tag{25}$$

Terminate:

$$P(O|\ \lambda)=\sum_{i=1}^{N}\alpha_T(i) \tag{26}$$

The first step is to set the forward probability to the combined probe of the state and the first observable. The main component of the algorithm is the second step. If the phantom is in any of the N potential ones at moment t, it will transfer to the state at moment t+1 with a specific probability. Step 3 The sum of all aT(i) is P(O|λ) according to the definition of forward probability.

Corresponding to the forward probability, the backward probability is defined as:

$$\beta_t(i)=P\left(O_{t+1}O_{t+2}\ldots O_T\middle|\ q_t=S_i,\lambda\right) \tag{27}$$

Initialization:

$$\beta_T(i)=1,\quad 1\le i\le N \tag{28}$$

Iteration:

$$\beta_t(i)=\sum_{j=1}^{N}a_{ij}b_j\left(O_{t+1}\right)\beta_{t+1}(j),\quad 1\le t\le T-1,1\le j\le N \tag{29}$$

Terminate:

$$P(O|\ \lambda)=\sum_{i=1}^{N}\beta_1(i) \tag{30}$$

By using forward and backward probabilities, the likely output of the entire observational series to the HMM model can be divided into the product of the probabilities of the two observational series. By using the corresponding recurrence formula, the following output probability calculation formula can be obtained:

$$P(O|\ \lambda)=\sum_{i=1}^{N}\alpha_t(i)\beta_t(i)=\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j\left(O_{t+1}\right)\beta_{t+1}(j),\quad 1\le t\le T-1 \tag{31}$$

The Viterbi algorithm is usually used, which is a dynamical engineering-based approach for searching for a singularly best sequence of states [20]. The problem to be solved by the Viterbi algorithm is to determine a sequence of states that maximizes the probability of the outcome given a given sequence of observations and a module, i.e., the problem of identification.

$$\delta_t(i)=\max_{q_1,q_2,\ldots,q_{t-1}}P\left(q_1,q_2,\ldots,q_t,q_t=S_i,O_1,O_2,\ldots O_t|\ \lambda\right) \tag{32}$$

If the state of the system at point t is Si, the optimal path is traced back to time T-1 along the optimal route, and a state labeling of the system is introduced, then the process of searching for the optimal succession of states is as follows:

Initialization:

$$\begin{gathered}\delta_1(i)=\pi_i b_i\left(O_1\right),1\le i\le N\\ \psi_1(i)=0\end{gathered} \tag{33}$$

Iterative calculation:

$$\begin{gathered}\delta_t(j)=\max_{1\le i\le N}\left|\delta_{t-1}(i)a_{ij}\right|b_j\left(O_t\right),\quad 2\le t\le T,1\le j\le N\\ \psi_t(j)=\arg\max_{1\le i\le N}\left|\delta_{t-1}(i)a_{ij}\right|,\quad 2\le t\le T,1\le j\le N\end{gathered} \tag{34}$$

Termination calculation:

$$\begin{gathered}p^*=\max_{1\le i\le N}\delta_T(i)\\ q_T^*=\arg\max_{1\le i\le N}\delta_T(i)\end{gathered} \tag{35}$$

Path backtracking:

$$q_t^*=\psi_{t+1}(i)\left(q_{t+1}^*\right),\quad t=T-1,T-2,\ldots,1 \tag{36}$$

A statistical model of voice signals' time series structure, known in mathematics as a double stochastic process, is called the hidden Markov model.

Two methods are available for simulating the change in the statistical properties of the speech signal: an implicit random process that employs a Markov chain with a finite number of states, and a random process that utilizes the observation sequence linked to each stage of the Markov chain. Although the latter expresses the previous, the former's actual boundaries are incalculable.

For voice recognition, there are four pattern-matching steps: feature extraction, template training, template classification, and judgment. Fig. 2 illustrates its fundamental framework:

Voice identification using HMM is essentially a probabilistic operation. After calculating the model parameters based on the training set dataset, it is also required to calculate the conditional probability (Viterbi's algorithm) of each model separately based on the test set data, and the model with the largest likelihood is the recognition result [21].

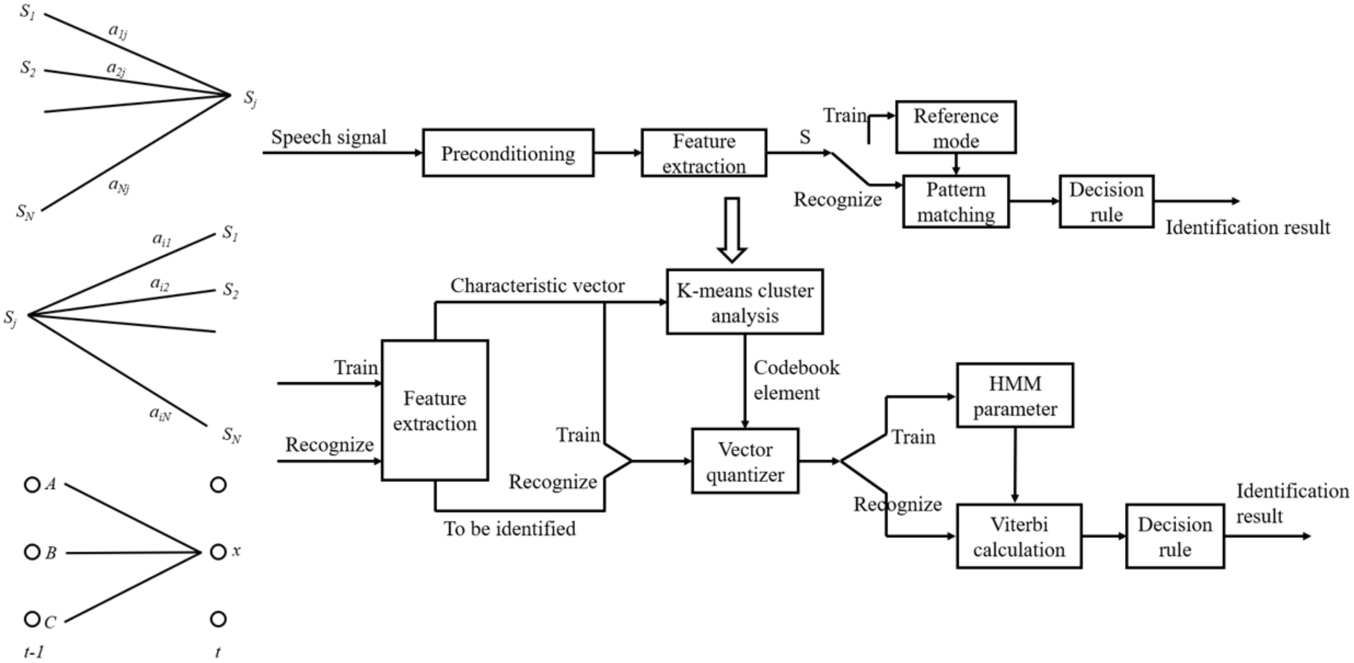


Fig. 2. Hidden markov model architecture.

Model parameter estimation is the process of training the pattern parameters, i.e., the series of known observations, and the output probability can be maximized by adjusting the model parameters. The specific steps are as follows: first, initialize the model parameters, then use some algorithm to input the same sequence of observations as the sequence of training samples, and repeatedly correct the sequence of observations, so that the pattern parameter with the maximum output potential becomes the final trained model parameter [22]. Since the Baum-Welch method is based on the maximum likelihood criterion, it has the advantages of fast convergence speed and monotonous growth of likelihood value, so the Baum-Welch method is usually used for parameter re-estimation.

The Baum-Welch algorithm uses the maximum likelihood criterion to obtain a fresh pair of variables by replacing the observed values and the initial model parameters. In other words, the results computed from multiple substitutions according to the re-estimation formula are more representative of the observed series than the original inputs. This process is repeated until convergence, which is the desired model parameter.

$$\xi_t(i,j) = p\left(q_t = i, q_{t+1} = j \middle| \; O, \lambda\right) \tag{37}$$

According to the forward-backward algorithm, we can get:

$$\xi_t(i,j) = \frac{p\left(q_t = i, q_{t+1} = j, O \middle| \; \lambda\right)}{P(O/\lambda)} = \frac{\alpha_t(i) a_{ij} b_j\left(O_{t+1}\right)\beta_{t+1}(j)}{\sum_{i=1}^{N}\alpha_t(i)\beta_t(i)} \tag{38}$$

Then, the odds that the observed sequence is in a certain condition at time t:

$$\gamma_t(i) = \sum_{j=1}^{N}\xi_t(i,j) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N}\alpha_t(i)\beta_t(i)} \tag{39}$$

The flowchart of the Baum-Welch algorithm is below in Fig. 3.

B. *Application of the HHM Model in Speech Recognition*

The choice of state type is the first problem to be considered when HMM is applied to speech recognition. There are two kinds of HMM model structure: each state traversal and left to right. The HMM model of state traversal can be applied to speaker recognition, language recognition, and so on. According to the characteristics of the human pronunciation process, the HMM of the right and left models is generally selected in speech recognition. In this paper, no span left-right model is adopted. There is no clear regulation on the number of states, which needs to be clarified through experiments and experience [23]. In English isolated word recognition technology, generally, the number of states between 4-8 is sufficient to achieve better results. Through the analysis of the test results, the amount of states N=4 is selected.

In addition to the number of states, the Gaussian mixture number of observed probability density functions also determines the final recognition rate. Through the analysis of subsequent experimental simulation results, the Gaussian mixture model number selected in this paper is M=3.

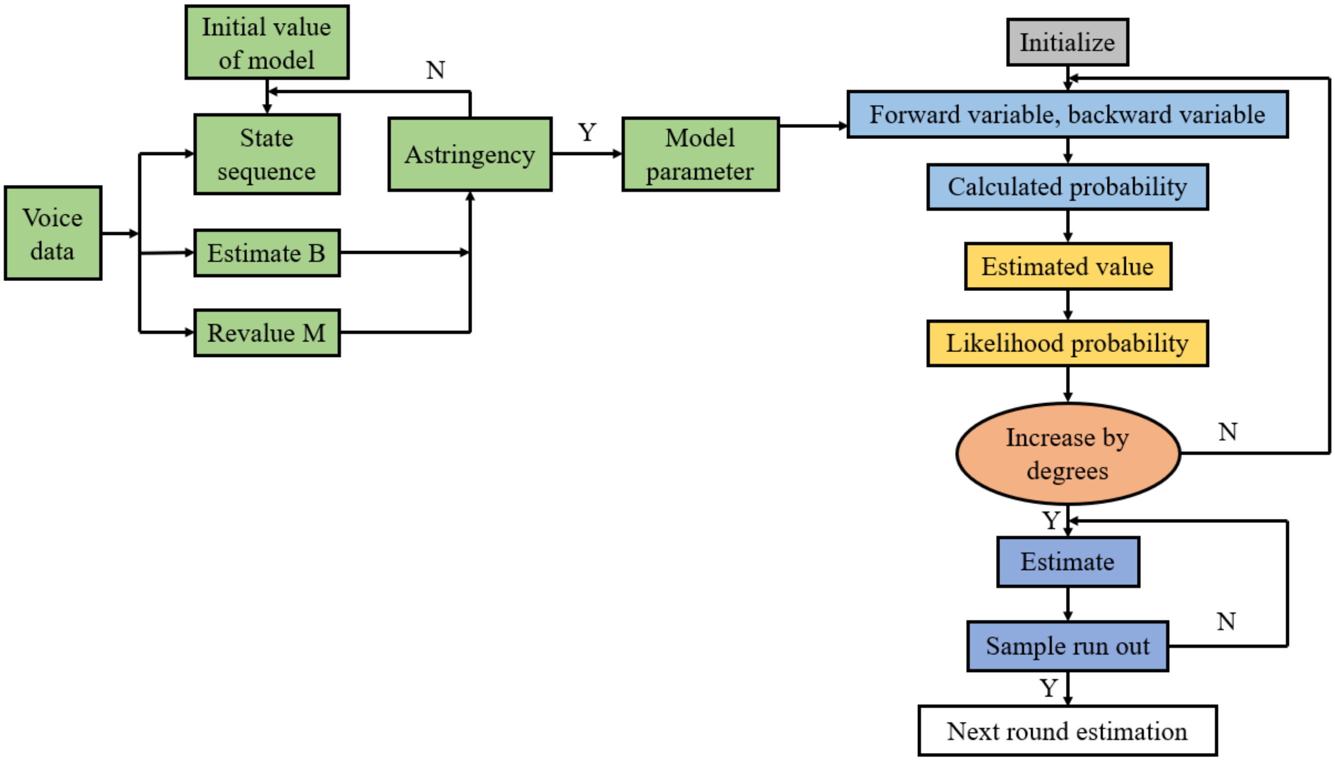


Fig. 3. Flowchart of Baum-Welch algorithm.

When applying the Baum-Welch algorithm to the parameter training of HMM models, how to correctly determine the initial parameters of the HMM to keep the partial maximization value as close as feasible to the global optimum is the focus of research [24]. In addition, a good choice of initial values can also reduce the number of iterations required for convergence, i.e., improve the computational efficiency. According to experience, the selection of the internal probability and initial value of the state transformation matrix does not have much influence on the recognition rate, and non-zero random numbers or uniform values can be used. Therefore, the initial probability of the state transition matrix and the choice of the initial value of the state transition matrix do not have much influence on the recognition rate, and non-zero random numbers or uniform values can be used. Based on this consideration, this paper adopts a composite clustering algorithm that combines the piecewise K-means algorithm, Viterbi algebras, and Baum-Welch algebras.

The basic idea of the method is as below:

1) Establish the initial parameters of the HMM model.

2) According to λ, use the Viterbi method to classify the input training speech database into the most probable state sequences.

3) Re-estimation of the input B using the "segmented K-means" method for the continuous HMM assumption using a hybrid Gaussian density function with M number of mixtures. The speech parameters corresponding to a particular state are pooled together and a K-means clustering operation is performed to classify the trained state speech clusters into M classes [25]. Then the meaning and covariance of the speech parameters of the same class are computed as the meaning vector and variance covariance mosaic of the class, thus obtaining the M normally distributed parameters of the M classes. Finally, the mixture weights of the class density functions are obtained by taking the number of speech frames included in each class and dividing it by the number of all speech in that state. This gives a new B, and thus a new set of initial values.

4) The λ obtained in step (3) is used as the initial value for the BW parameter re-estimation method to perform parameter re-estimation on the HMM module, thus obtaining the new module variables.

5) If the difference is less than the preset min value, it indicates that the model parameters have converged, there is no need to re-estimate, and λ is the final parameter output. Otherwise, λ continues the iteration as the new initial argument.

Since the segmented K-means algorithm is the idea of state optimization to carry out the maximum likelihood criterion, it can greatly accelerate the training speed of the model by realizing the initial parameter re-estimation.

IV. Optimize the DNN-HMM Model and its Analysis

A. Deep Neural Network

A neural network is also composed of neurons as shown in Fig. 4, which is a simple nervous web. The net is composed of an input level, a hidden level, and an output level. The nodes directly connected between layers are fully connected. This network model can fit simple nonlinear transformations.

When the amount of input data increases, it is necessary to fit complex nonlinear transformations, which can be achieved by adding the number of neurons in the hidden layer or the number of layers in the hidden layer. As shown in Fig. 4, the number of neurons and the number of hidden layers in a deep Neural Web increases accordingly.

The network can be trained using the error backpropagation algorithm, which can accommodate more complex functions. To meet the large increase in the amount of data, the number of layers of the deep neural network needs to be increased. At this point, the amount of network layers increases, the parameters surge, and training becomes more difficult, and the results are often difficult to converge.

The activation function in neurons is nonlinear, and the input is digital information, through which certain mathematical operations can be performed [26]. According to different excitation forms, there are different activation functions:

Sigmoid activation function:

$$f(x) = \frac{1}{1+\exp(-x)} \tag{40}$$

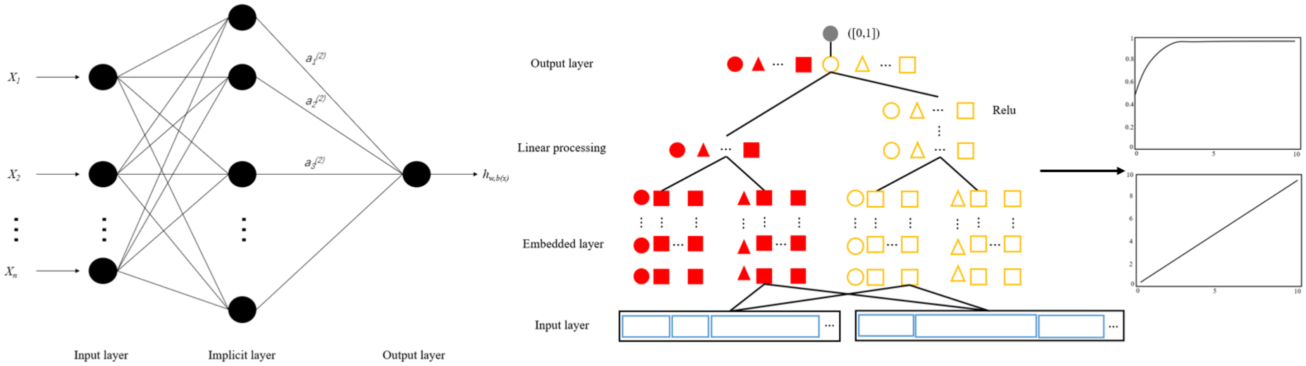


Fig. 4. Deep neural network model.

Tanh activity feature:

$$f(x) = \tanh(x) \quad (41)$$

ReLU activity feature:

$$f(x) = \max(x, 0) \quad (42)$$

The featured graph of the Tanh function compensates the output value between [-1, 1] and suffers from the same saturation problem as the sigmoid feature, but its output is zero-centered. Therefore, in practice, the Tanh nonlinear function is widely used.

Compared with the sigmoid and tanh functions, the ReLU function has a great acceleration effect on the convergence of stochastic gradient descent.

B. Optimization of DNN-HMM Model

The DNN-HMM model in speech recognition combines HMM and DNN, each performing its duties and sharing different tasks. HMM models speech timing signals. DNN models the posterior probability distribution of the input sample.

The input of DNN is the feature vector extracted from each frame of a speech signal after it is divided into frames. However, the output vector is the probability of the corresponding HMM state, so its dimension is equal to the number of states in the HMM. But since this is a supervised learning task, the input-to-output mapping must be found. Although the speech information and its corresponding text information of each sample can be known from the training sample, it seems that this is the mapping relationship between the input speech and the output text, but the DNN-HMM does not directly model the whole speech, but the model of each frame signal. Therefore, it is necessary to obtain the HMM state label corresponding to each input vector in the DNN, and then use this label to train a DNN model. In order to obtain the HMM state labels corresponding to each utterance, a traditional GMM-HMM model is usually trained in advance. GMM-HMM is used to force align the training samples. Each signal in the observed sequence is aligned with the state of its corresponding HMM. After alignment, DNN can be used instead of GMM to calculate the observation probability in HMM for training.

The input to the model is the feature of successive frames, that is, in addition to the current frame, the information of previous and subsequent frames is included. In general, if the current frame time is t, the input feature also includes all information from time T-4 to time t+4. The combination of 9 consecutive features to represent the current situation effectively utilizes contextual information, which is one of the reasons that DNN is superior to GMM. However, for continuous speech recognition tasks, it is not enough to only use the context information of the frame to model the speech. It is necessary to take into account the long-term dependence of the speech signal in time. The model works better if it can model the long-term dependence of the speech signal using its historical information. However, DNN-HMM based on a feedforward neural network cannot do this.

TDNN also belongs to the feed forward neural network, and its network structure is shown in Fig. 5. Different from DNN, it is not fully connected between layers, but the connection is controlled by the delay parameter. For example, [-2,2] in hidden layer 1 represents that the current frame is taken as the basis, and the two frames before and after each frame are a total of five as the input of the network, and each neuron in the time domain connects with the previous layer according to this rule. The architecture of the hidden levels is the same, so the income of each hidden level in TDNN is not just the output of the preceding level at the present moment, but also the outcome of the preceding level at t moments around the previous level. This allows each hidden layer to extend the time domain, especially deeper into the network, which contains more information from the input layer.

From a local point of view, the same layer of TDNN can be divided into many repeated network structures. For example, the first hidden layer can be split into substructures as shown in Fig. 5.

C. Neural Network Model Training under Kaldi

DNN-HMM model training is supervised training, which means that alignment information is obtained by the GMM-HMM model before training. The regularity of input features is conducive to the training of model parameters, so the input features are transformed in different stages of the GMM-HMM model training. The total transformation process is shown in Fig. 6.

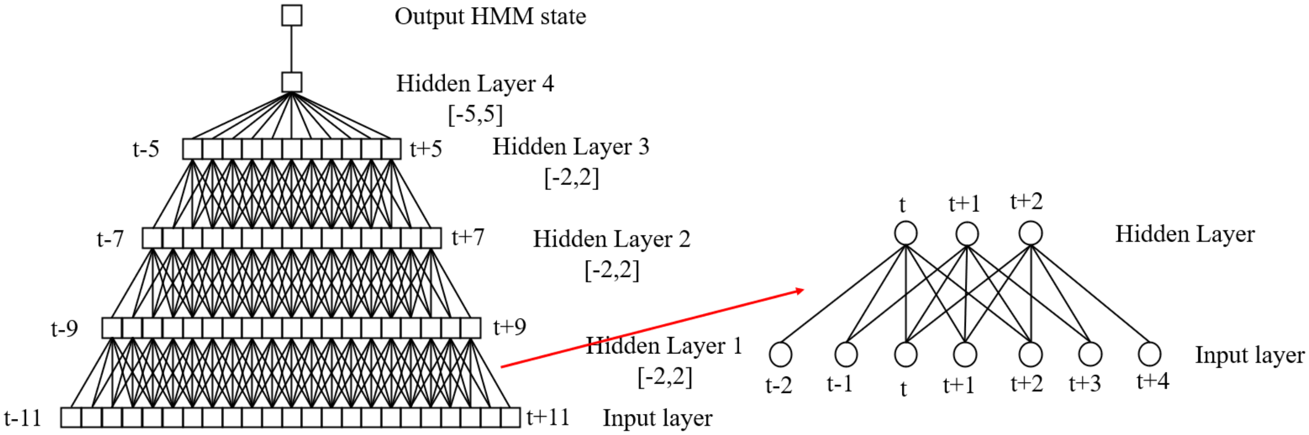


Fig. 5. TDNN structure diagram.

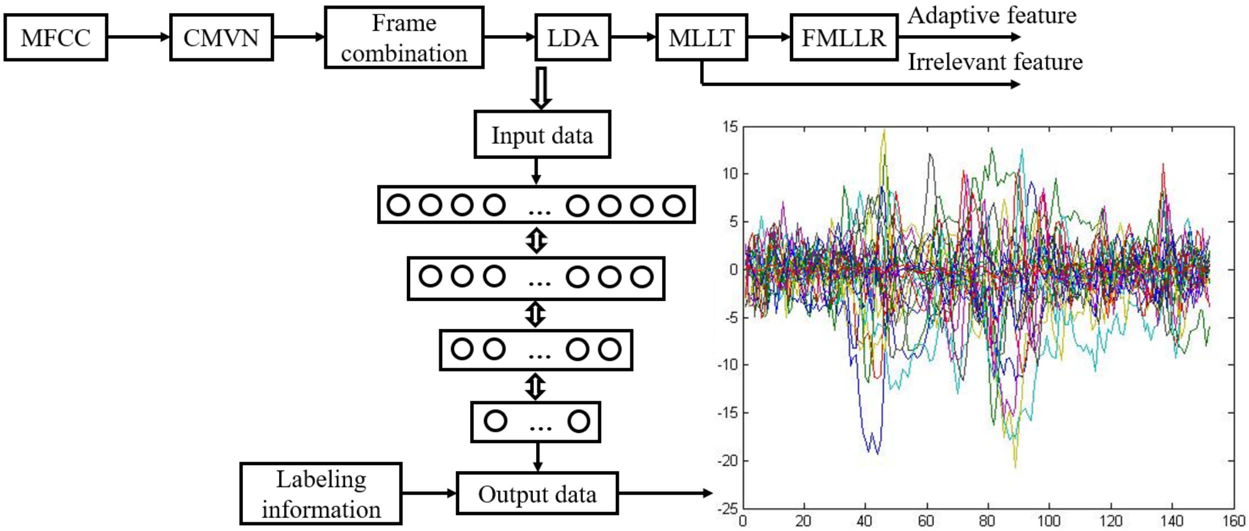


Fig. 6. DNN-HMM training process.

The GMM input feature has not been normalized after a series of processing, so another normalization process is needed. Based on this, several frames of data are spliced together forward and backward in the time domain to augment the modeling capability of the neural net, and the feature data need to be normalized and de-correlated again. Because the feature dimensions of time domain splicing are strongly correlated, it is not conducive to the training of neural networks. In this paper, CMVN is used to normalize the mean and variance of each dimension to achieve the purpose of de-correlation.

where, each element of the weight matrix specifies the weight of the edge between the concealed level cell and the visual level cell, with a bias for each visible level cell and a bias for each concealed level cell. Depending on the distribution of the random variables, the "energy" of the RBM is defined in two ways:

$$E(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j h_j w_{i,j} v_i \tag{43}$$

$$E(v,h) = \sum_i \left(a_i - v_i\right)^2 - \sum_j b_j h_j - \sum_i \sum_j h_j w_{i,j} v_i \tag{44}$$

Eq. (43) represents the equivalent energy function for the Bernoulli distribution and Eq. (44) represents the equivalent energy function for the Gaussian distribution. According to the Gibbs distribution, the probability of RBM in the current state is given by Eq. (45).

$$P(v,h) = \frac{1}{Z} e^{-E(v,h)} \tag{45}$$

$$Z = \sum_v \sum_h e^{-E(v,h)} \tag{46}$$

This probability can be considered as a joint probability distribution of the deep-level state and the hidden-level state, which is obtained by the RBM energy of the current state being normalized by the RBM energy of all possible states according to the exponential rule, where, Z is the partition function, which is a regular term that takes into account the energy of RBM in all states. Therefore, the edge distribution of the state vector of the display layer could be derived from the above joint distribution.

$$P(v) = \frac{1}{Z} \sum_h e^{-E(v,h)} \tag{47}$$

After Kaldi calls steps/net/pre-train dbn. sh for pre-training, it is time to train the neural network. The Truncated BPTT algorithm is usually used, that is, BPTT is carried out in a data block, which can contain a sentence or a fragment of fixed length. In Kaldi, steps/nnet/train.sh is called for training, the number of input layer units is set to 40, the number of hidden layers is fixed to 4, and the number of units in each level is set to 1024. the initial studying ratio is defined to be 0.008. the results obtained from the model's alignment will be used for the subsequent discriminative training.

D. Result and Discussion

According to the implementation method of the HMM model, PNN model, and DNN-HMM mixed model, MATLAB programming is carried out for the three models, and the three recognition systems are tested and analyzed with the help of the MATLAB simulation experiment platform.

All experiments were programmed using MATLAB for the three models, and the recognition system model program was split into two parts: the speech trainer and the speech recognition program. The speech test samples were 10 place names, namely Beijing, Shanghai, Guangzhou, Tianjin, Chongqing, Wuhan, Shenyang, Dalian, Changchun, and Harbin. The speakers were five men and five women, each pronouncing each word five times. The top 250 samples were taken as test samples and the bottom 250 samples were taken as training samples. The speech training program first builds a reference model library for these 10 words, and then the speech recognition program recognizes the results by comparing the input parameters with the reference model library. To validate the anti-jamming performance of the DNN-HMM mixture module, the HMM model and the PNN module are compared in a pure speech environment and a signal-to-noise environment, respectively. Table I displays the word recognition rate compared between the models in the pure speech environment. The first digit in the table is the number of correctly recognized words, and the last digit is the overall number of recognized letters. 24/25 means that 25 samples are input to be recognized, and the number of correctly recognized words is 24.

According to the speech recognition rate data in the above table, the test of this model in Beijing, Tianjin, Wuhan, Changchun and Harbin has reached 100%, while the test of this model in other regions is also close to 98%. The recognition rate of the optimized DNN-HMM mixed model is the highest, reaching 97.5%, followed by the HMM model with 95.4%, and the PNN model with 90.1% is the lowest. The data comparison shows that the hybrid model has better recognition performance than the single model, and the recognition rate is improved, which achieves the expected effect. There is little difference between the HMM and the optimized DNN-HMM modules in terms of recognition rate. To further illustrate the benefits and performance of hybrid models in speech recognition, we analyzed the anti-interference performance of each model. Table II displays a comparative identification ratio after adding different signal-to-noise ratios to each model.

From Tables I and II, it can be observed that as the SNR gradually decreases, the identification rate of the speech identification system continues to decrease, and the identification rates of the HMM model and the PNN model decrease obviously, however, the identification rate of the hybrid module decreases less obviously than that of the single model. The optimized DNN-HMM hybrid module incorporates the strong sequential handling capability of the HMM model and the excellent classifying capability of the probabilistic neural net. This combined ability can characterize the linguistic content of voice in a more comprehensive and detailed way, and improve the recognition rate, anti-interference performance, and ruggedness of the speech recognition system.

TABLE I. COMPARISON OF RECOGNITION RATES OF DIFFERENT RECOGNITION MODELS

Model ***Recognizer***	***HMM model***	***PNN model***	***Optimized DNN-HMM***
Peking	24/25	22/25	25/25
Shanghai	24/25	21/25	24/25
Guangzhou	21/25	22/25	24/25
Tianjin	25/25	23/25	25/25
Chongqing	24/25	22/25	24/25
Wuhan	23/25	24/25	25/25
Shenyang	21/25	22/25	24/25
Dalian	23/25	22/25	23/25
Changchun	25/25	21/25	25/25
Harbin	24/25	22/25	25/25
Product recognition rate	95.4%	90.1%	97.5%

TABLE II. COMPARISON OF RECOGNITION RATES OF DIFFERENT SNRS OF DIFFERENT MODELS

model ***Signal-to-noise ratio***	***HMM model (%)***	***PNN model (%)***	***Optimized DNN-HMM (%)***
5db	31.3	45.3	64.1
10db	63.4	70.3	79.9
15db	75.1	77.3	89.1
20db	88.1	86.5	93.6
25db	89.8	89.3	94.7
30db	93.6	90.4	96.7

V. CONCLUSION

In this paper, a web-based oral English teaching system is proposed by combining speech recognition technology with a deep neural network-Markov model. The continuous speech signal is modeled by combining DNN and HMM. To solve the long-term dependence problem of DNN, TDNN is used to reconstruct the acoustic model, its structure and implementation principle are introduced, and the neural network model is built and trained by Kaldi. Specific conclusions are as follows:

First, this paper takes the technical principles of speech identification as the theoretical basis for a thorough study and research. Speech identification system according to the Hidden Markov Model. In speech processing, MFCC is used as the feature principle of speech, and the Hidden Markov Model is utilized for training and recognition, to produce a speech recognition system.

Second, because GMM-HMM does not utilize the context information of frames, DNN is introduced to construct a new acoustic model DNN-HMM, and its structure and training algorithm are introduced. However, DNN does not take into account the long-term dependence on the time of speech signals, so TDNN is used to reconstruct the acoustic model to introduce its structure and advantages, build and train the neural network based on Kaldi, and apply the DT method to train the neural network model to continuously improve the performance of the acoustic model.

Third, the simulation results show that in terms of speech identification rate, the optimized DNM-HMM hybrid model has the highest recognition rate of 97.5%, followed by the HMM model with 95.4%, and the PNN model has the lowest identification rate of 90.1%. The optimized DNN-HMM hybrid model achieves a high recognition rate of 96.7% when the signal-to-noise ratio is 30db. It incorporates the strong sequential processing capability of the HMM model and the excellent categorization capability of the probabilistic neural network, which can characterize the semantic meaning of speech in a more comprehensive and detailed way.

REFERENCES

[1] Abdel-Hamid O, Mohamed A, Jiang H, et al. Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and language processing, 2014, 22(10): 1533-1545. Doi: 10.1109/TASLP2014.2339736.

[2] Kamble B C. Speech recognition using an artificial neural network–a review. Int. J. Comput. Commun. Instrum. Eng, 2016, 3(1): 61-64.

[3] Zhang Jing, Yang Jian, Su Peng. A review of monosyllabic recognition in Speech recognition. Computer Science, 2019, 47(S2):172-174+203. (in Chinese).

[4] Yang Yang, Wang Yuduo. Speech recognition based on improved Convolutional neural networks. Applied Acoustics, 2018, 37(06):940-946. (in Chinese).

[5] Sun Jian, Guo Wu. Japanese Speech recognition based on Link sequential classification. Minicomputer Systems, 2018, 39(10):2129-2133. (in Chinese).

[6] Huang Yulei, Luo Xiaoxia, Liu Duren. MFSC coefficient characteristics of locally finite weighted sharing CNN speech recognition. Journal of control engineering, 2017, 24 (7): 1507-1513.

[7] Hou Yimin, Zhou Huiqiong, Wang Zhengyi. Application Research of Computers, 2017, 34(08):2241-2246.

[8] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks. International conference on machine learning. PMLR, 2014: 1764-1772.

[9] Amberkar A, Awasarmol P, Deshmukh G, et al. Speech recognition using recurrent neural networks. 2018 international conference on current trends towards converging technologies (ICCTCT). IEEE, 2018: 1-4. Doi: 10.1109/ICCTCT.2018.8551185.

[10] Nassif A B, Shahin I, Attili I, et al. Speech recognition using deep neural networks: A systematic review. IEEE access, 2019, 7: 19143-19165. Doi:10.1109/ACCESS.2019.2896880.

[11] Dua S, Kumar S S, Albagory Y, et al. Developing a Speech Recognition System for Recognizing Tonal Speech Signals Using a Convolutional Neural Network[J]. Applied Sciences, 2022, 12(12): 6223. https://doi.org/10.3390/app12126223.

[12] Swietojanski P, Ghoshal A, Renals S. Convolutional neural networks for distant speech recognition. IEEE Signal Processing Letters, 2014, 21(9): Doi:1120-1124. 10.1109/LSP.2014.2325781.

[13] Lokesh S, Malarvizhi Kumar P, Ramya Devi M, et al. An automatic tamil speech recognition system by using bidirectional recurrent neural network with self-organizing map[J]. Neural Computing and Applications, 2019, 31: 1521-1531. https://doi.org/10.1007/s00521-022-08144-x.

[14] Islam J, Mubassira M, Islam M R, et al. A speech recognition system for Bengali language using recurrent neural network. 2019 IEEE 4th international conference on computer and communication systems (ICCCS). IEEE, 2019: 73-76. Doi: 10.1109/CCOMS.2019.8821629.

[15] Fohr D, Mella O, Illina I. New paradigm in speech recognition: deep neural networks. IEEE international conference on information systems and economic intelligence. 2017.

[16] Vydana H K, Vuppala A K. Residual neural networks for speech recognition. 2017 25th European Signal Processing Conference (EUSIPCO). IEEE, 2017: 543-547. Doi: 10.23919/EUSIPCO.2017.8081266.

[17] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks. 2013 IEEE international conference on acoustics, speech and signal processing. Ieee, 2013: 6645-6649. Doi:10.1109/ICASSP.2013.6638947.

[18] Siniscalchi S M, Yu D, Deng L, et al. Exploiting deep neural networks for detection-based speech recognition. Neurocomputing, 2013, 106: 148-157. https://doi.org/10.1016/j.neucom.2012.11.008.

[19] Rani P, Kakkar S, Rani S. Speech recognition using neural network. International journal of computer applications, 2015, 4: 11-14.

[20] Waibel A. Modular construction of time-delay neural networks for speech recognition. Neural computation, 1989, 1(1): 39-46. https://doi.org/10.1162/neco.1989.1.1.39.

[21] Song W, Cai J. End-to-end deep neural network for automatic speech recognition. Standford CS224D Reports, 2015: 1-8.

[22] Sainath T N, Weiss R J, Wilson K W, et al. Multichannel signal processing with deep neural networks for automatic speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(5): 965-979. Doi:10.1109/TASLP.2017.2672401.

[23] Mustafa M K, Allen T, Appiah K. A comparative review of dynamic neural networks and hidden Markov model methods for mobile on-device speech recognition[J]. Neural Computing and Applications, 2019, 31: 891-899. https://doi.org/10.1007/s00521-017-3028-2.

[24] Chan W, Jaitly N, Le Q, et al. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2016: 4960-4964. Doi: 10.1109/ICASSP.2016.7472621.

[25] Al Smadi T, Al Issa H A, Trad E, et al. Artificial intelligence for speech recognition based on neural networks. Journal of Signal and Information Processing, 2015, 6(02): Doi:66. 10.4236/jsip.2015.62006.

[26] Saksamudre S K, Shrishrimal P P, Deshmukh R R. A review on different approaches for speech recognition system. International Journal of Computer Applications, 2015, 115(22).

Recurrence Prediction and Risk Classification of COPD Patients Based on Machine Learning

Xin Qi[1], Hong Chen[2]*
Academic Affairs Office, Heilongjiang University of Chinese Medicine, Harbin 150001, China[1]
Chinese Pediatrics, First Affiliated Hospital, Heilongjiang University of Chinese Medicine, Harbin 150040, China[2]

***Abstract*—In response to the frequent recurrence and readmission of patients with chronic obstructive pulmonary disease, a machine learning based recurrence risk prediction and risk classification model for patients with chronic obstructive pulmonary disease is studied and constructed. Approach: This model first utilizes the optimized long short-term memory network to recognize named entities in patient electronic medical records and extract entity features. Then, XGBoost is used to predict the probability of patient relapse and readmission, and its risk is classified. Results: These results confirm that the optimized bidirectional long short-term memory network has the best performance with an accuracy of 84.36% in electronic medical record named entity recognition. The accuracy of XGBoost is the highest on both the training and testing sets, with values of 0.8827 and 0.8514, respectively. XGBoost has the best predictive ability and effectiveness. By using k-means for layering, the workload of manual evaluation was reduced by 91%, and the overall simulation accuracy of the model was as high as 97.3% and 96.4%. Conclusions: These indicate that this method can be used to balance high-risk patients between risk, cost, and resources.**

Keywords*—Machine learning; COPD; BiLSTM; XGBoost; k-means; recurrence; risk classification

I. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a type of lung disease characterized by airflow restriction, which is a long-term, irreversible, and progressive chronic respiratory disease [1]. COPD is a common disease with a high incidence rate, which poses a heavy burden on the physical health and disease burden of the people [2]. It has become an important public health problem that restricts China's economic and social development [3]. Due to factors such as medical quality, hygiene effectiveness, and economy, some patients may relapse and be hospitalized within a month due to the same reasons, which has become a major challenge facing the world [4-5]. Predicting preventable frequent relapse hospitalization and understanding the causes of relapse hospitalization are currently important topics of widespread concern [6]. Machine Learning (ML) is a scientific method that enables prediction, identification, and decision-making of environmental changes without human intervention [7]. ML has achieved rapid development in medicine, especially in early warning of diseases, providing convenient and fast decision support for people [8]. Currently, ML and data mining have become methods that can potentially improve the predictive ability of relapse admission risk prediction models [9]. How to use ML technology to predict the recurrence of COPD patients and classify their risk is the main problem of this study. Currently, research on recurrence prediction and risk classification for COPD patients mainly relies on the experience and professional knowledge of clinical doctors, lacking objective and systematic prediction models. Therefore, the research on using ML technology for predicting the recurrence and risk classification of COPD patients still needs to be improved. The main objective of this study is to use ML technology to establish a model that can accurately predict the recurrence of COPD patients and classify their risk, improve the management and treatment level of COPD patients, reduce the risk of recurrence, and improve their quality of life. The innovation of this study lies in the full utilization of electronic medical record text information, while also ensuring the personal information of patients.

The article consists of four sections. In Section I, there is a literature review that introduces the relevant research content of different scholars. Section II outlines the related works. Section III is the research method, which mainly introduces the Named Entity Recognition (NER) of electronic medical records of COPD patients, as well as the prediction and risk classification of relapse and readmission of COPD patients. Section IV is the result analysis, which explains the electronic medical record NER of COPD patients under different ML algorithms, as well as the results of predicting relapse, readmission, and risk classification of COPD patients. Discussion is presented in Section V. Finally, Section VI is the conclusion that summarizes the results of recurrence prediction and risk classification for COPD patients and points out the shortcomings of the research.

II. Related Works

The Bidirectional Long Short-Term Memory Network (BiLSTM) is widely used in NER and predictive analysis, and many scholars have achieved good research results. To improve the NER efficiency of product review, researchers such as Postiche H proposed a product review NER method based on BiLSTM. These experiments confirm that the research method is more efficient in identifying named entities in product reviews compared to current methods. Benali B A et al. proposed a multi-head self-attention mechanism based on structures such as BiLSTM to improve the accuracy of NER in natural language texts, to perform NER on natural language in social media. These experiments confirm that this method can combine characters and words in the embedding layer, resulting in significantly better recognition results than current naming recognition methods [10]. Long R and other scholars proposed a disease diagnosis depth framework that integrated BiLSTM to enhance the analysis of electronic medical record data. These experiments confirm that the framework can significantly improve the performance of disease diagnosis [11]. Puh K et al. designed a deep learning model based on BiLSTM to predict the emotions in tourist comments in the tourism industry to extract and rate the emotions in tourist comments. These experiments confirm that this deep learning model has more efficient work efficiency and more accurate prediction results compared to other models [12].

ML has become an emerging and effective method for predicting the risk of disease recurrence. Matheson A M and other researchers proposed an ML-based prediction method to address the high risk of COPD recurrence. These experiments confirm that this method's predicting accuracy is superior to traditional regression analysis methods, which is beneficial for the prevention of COPD recurrence [13]. AL khadar and other scholars designed a recurrence prediction model based on decision tree classifier to enhance the prediction efficiency of recurrence and survival rate in oral cancer patients. These experiments confirm that the research method has the best predictive performance compared to other traditional ML models [14]. To compare the predictive performance of ML algorithm for early biochemical recurrence after prostatectomy, Wong NC et al. selected three supervised ML algorithms to construct models for prediction and compared them with traditional regression analysis. These experiments confirm that these three ML models have high accuracy in predicting biochemical recurrence [15]. Paredes AZ and other scholars proposed a fusion ML index prediction method to address the issue of high postoperative recurrence rates in patients with liver metastasis from colorectal cancer. This method combines resampling method with multivariate mixing effect to construct a prediction model. These experiments confirm that this method has a high accuracy in predicting indicators, which is beneficial for clinical analysis of postoperative recurrence risk [16].

In summary, BiLSTM can improve the accuracy of NER, and ML outperforms traditional analysis methods in predicting the risk of disease recurrence. However, the above studies lack the application of NER in predicting the recurrence risk of disease patients. There are few studies that combine these two and use ML to predict the risk of disease recurrence based on NER of electronic medical records. Therefore, to predict and classify the risk of recurrence in COPD patients, this study uses ML to perform NER on the patient's electronic medical record and predicts the re-admission of recurrence, providing a certain scientific basis for medical institutions.

III. ML Based Recurrence Prediction and Risk Classification Model for COPD Patients

A recurrence prediction and risk classification model for COPD patients is constructed for the prediction of recurrence. Firstly, BiLSTM-CRF is used to perform NER on electronic medical records. Then, XGBoost is used to predict the recurrence and readmission of COPD patients. Finally, k-means is used to classify the recurrence risk of COPD patients, as well as early prevention and effective intervention.

A. Electronic Medical Record NER for COPD Patients Based on Optimized BiLSTM

The electronic medical record collects rich diagnosis and treatment information from patients, including detailed historical data of patients seeking medical treatment and treatment in the hospital. It is the foundation and core of medical and health big data. There are various forms of electronic medical record storage, including structured and textual data. If people only focus on easily processed structured data and ignore textual data, it will lead to the inability to obtain certain characteristics of patient real data [17]. When designing a method for predicting the recurrence and readmission of COPD patients, it is first necessary to extract medical entities with more features from text data in electronic medical records and structurally process them. Fig. 1 shows the NER of an electronic medical record.

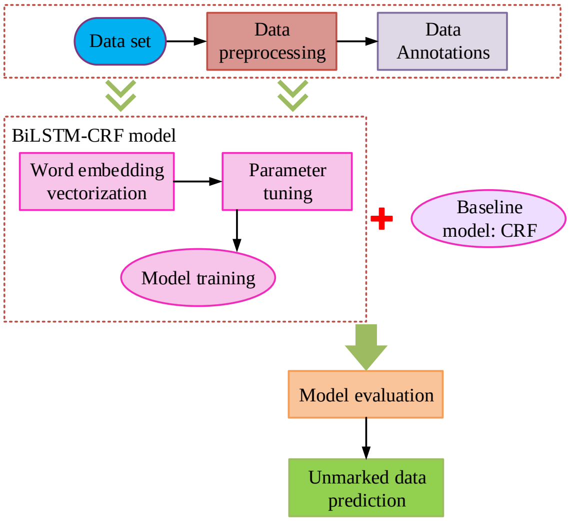


Fig. 1. The recognition process of named entities in electronic medical records.

According to Fig. 1, in electronic medical records' NER, it should first obtain data, and then preprocess and annotate the data. In the experiment, an electronic medical record NER method based on BiLSTM-CRF is designed, followed by word embedding vectorization, parameter tuning, and model training. Due to the fact that CRF is a commonly used method superior to other methods, it is used as a baseline method for method evaluation and prediction of unlabeled data.

The electronic medical record data in this study are sourced from the hospitalization records of patients in a tertiary hospital in Zaozhuang City. COPD patients who were hospitalized in the hospital from January 2020 to January 2021 were selected. With consent, the researchers logged into hospital's patient information management system and searched for target patient information. And it was deprivileged to obtain the inpatient information table and inpatient case text. Among them, textual information written by doctors, such as admission and exit records, can promote the design of predictive methods for patient relapse and readmission, and needs to be processed using the NER method based on ML. Before performing NER on inpatient electronic medical records, it is necessary to standardize the electronic medical records, including desensitization of patient data, detection and processing of sick sentences, and removal of irrelevant symbols [18]. Useful entity information was extracted for predicting the recurrence and readmission of COPD patients, mainly divided into comorbidities, examination indicators, indicator values, disease course, past physical fitness, and lifestyle habits. Afterwards, the data entities were annotated to construct the entity annotation dataset for the training model.

When designing the NER method based on BiLSTM, it is necessary to first perform word vector embedding transformation. Word embedding can solve the problem of vector encoded word similarity by converting text into numerical values. On this basis, the BiLSTM-CRF system is used to identify entities defined in the electronic medical records of hospitalized COPD patients. This system mainly includes two parts: BiLSTM layer and CRF loss layer. The former is essentially a traditional recurrent neural network. Due to its risk of gradient vanishing or exploding, it utilizes Long and Short-Term Memory (LSTM) for long-distance dependency capture [19]. Since LSTM cannot encode information from back to front, it is optimized to obtain BiLSTM to capture bidirectional semantic dependencies. The outputs of BiLSTM are all independently selected markers with higher scores, and there is no significant correlation between each marker. CRF can extract a certain number of constraint rules from samples and adjust them at the syntactic level. Therefore, this can enhance the constraints on samples, reduce the possibility of illegal sequences appearing, improve the prediction accuracy of labeled sequences, and ensure the generation of globally optimal labeled sequences. A BiLSTM-CRF system was constructed by combining BiLSTM and CRF. BiLSTM was used to learn features in the training set, including sub embedding and other features. After feedback to CRF layer, the tag sequence with the highest score was output. For a certain input sentence $[x]_1^T$, $[f\theta]i,t$ means the score of the i-th tag at time t. $[A]_{i,j}$ represents the probability of the ij-th position in the transition probability matrix of CRF, that is, the probability of transitioning from state i to state j. So for a certain input sentence and its label sequence $[i]_1^T$, Eq. (1) is the final score.

$$S\left([x]_1^T,[i]_1^T,\theta\right)=\sum_{t=1}^{T}\left([A]_{[i]_{t-1},[i]_t}+[f\theta]_{[i]_t,t}\right) \tag{1}$$

In Eq. (1), S represents the final score. A BiLSTM-CRF system has been successfully constructed to recognize entity information in electronic medical record texts. This model mainly consists of a word vector, BiLSTM, and CRF layers in Fig. 2.

Through Fig. 2, the system takes the text sequence of medical records as input and the relevant entities in the medical records as output. In the implementation of the system, each character input is first converted into a vector form through Word2Vec, which is used as the input of BiLSTM to extract contextual features. The output feature vector is used as the input of the CRF layer, and the input is normalized and the final label sequence is output. BiLSTM can provide users with more complete contextual information, thereby better understanding contextual dependencies. On this basis, an additional CRF layer has been added to further improve the existing BiLSTM system. By comprehensively optimizing the recognition results at the sentence level, the shortcomings of BiLSTM system are compensated.

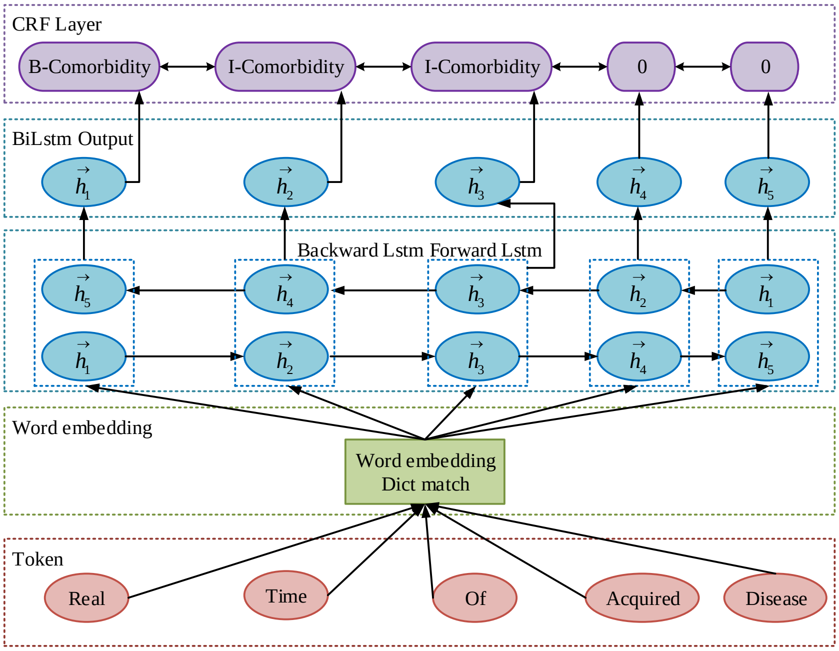


Fig. 2. BiLSTM-CRF network architecture.

B. Prediction of Recurrence in COPD Patients Based on XGBoost

This model's second layer is the prediction of relapse and readmission. That is, based on the medical information generated by the patient during hospitalization, it determines whether they will relapse and be re-admitted due to their physical condition for a period of time after discharge. 30 days are usually the most commonly used readmission threshold. Therefore, in this study, the readmission threshold is set to 30 days when predicting the recurrence and readmission of COPD patients [20]. With medical big data accumulating and ML progressing continuously, the prediction and prediction technology for relapse and readmission can effectively identify potential risk factors, providing a basis for the diagnosis and treatment of relapse and readmission patients. Fig. 3 shows the ML-based prediction of relapse and readmission in COPD patients.

According to Fig. 3, in the ML-based prediction of relapse and readmission of COPD patients, it is necessary to first integrate geometric features and then clean the data. Afterwards, a recurrence and readmission prediction system for COPD patients is designed and evaluated. Based on the previous section of NER, the extracted entity information and features are preserved. Due to the presence of noise and missing raw data, it is necessary to preprocess it to carry out predictions and obtain accurate prediction results. Fig. 4 shows the preprocessing content.

According to Fig. 4, preprocessing includes feature processing for comorbidities, assignment of discrete features, handling of outliers and missing values, descriptive statistics of variables, and handling of classification imbalance. Using ML algorithm to construct a relapse readmission prediction system for COPD patients, considering the system applicability, it should find the optimal classification method to achieve the best prediction effect. XGBoost is similar to a decision tree, in which each tree is divided by a special interval in its columns. On each tree shape, special attention should be paid to the sample with the highest error rate on the previous tree shape, so that there will be more optimization samples on the next tree shape. Finally, the results of each tree are combined and the weighted average method is generally used to obtain the results of each tree [21]. When adjusting XGBoost parameters, it should first adjust the learning rate of the system, followed by the total number of estimators used, the depth of each estimator, and finally the L1 and L2 regularization parameters [22]. To evaluate the classification performance, it should use five indicators: accuracy, accuracy, recall, F1 value, and ROC. Eq. (2) represents the accuracy.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

In Eq. (2), ACC represents accuracy. TP, FN, TN, and FP represent positive cases that are correctly classified, positive cases that are misclassified as negative cases, negative cases that are correctly classified, and negative cases that are misclassified as positive cases, respectively. Eq. (3) represents the precision.

$$P = \frac{TP}{TP + FP} \tag{3}$$

In Eq. (3), P represents precision. Eq. (4) represents the recall rate.

$$R = \frac{TP}{TP + FN} \tag{4}$$

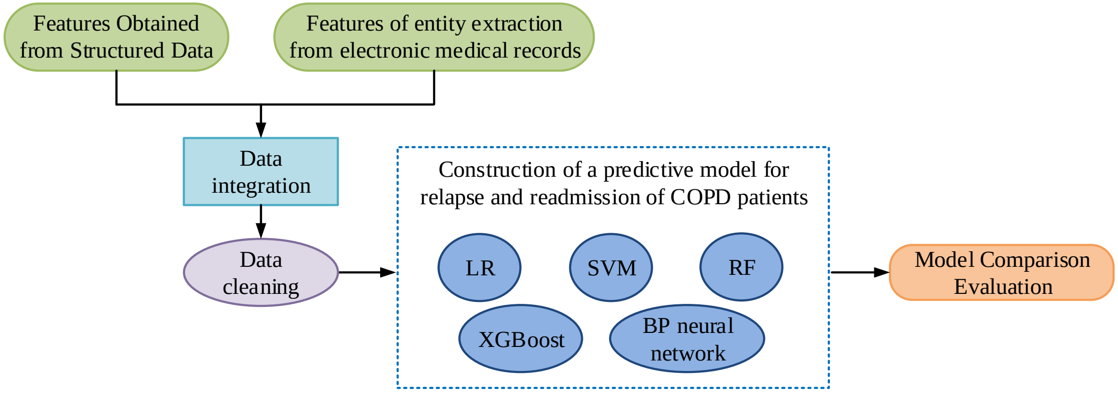

Features Obtained from Structured Data
Features of entity extraction from electronic medical records
Data integration
Data cleaning
Construction of a predictive model for relapse and readmission of COPD patients
LR
SVM
RF
XGBoost
BP neural network
Model Comparison Evaluation

Fig. 3. Machine learning based prediction process for relapse and readmission of COPD patients.

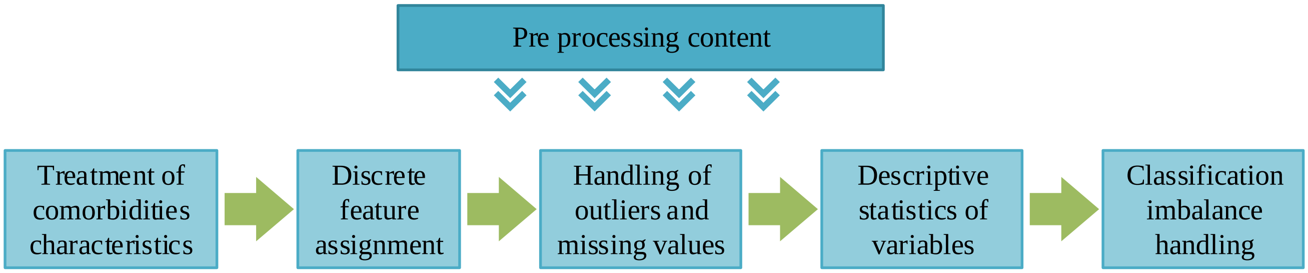

Pre processing content
Treatment of comorbidities characteristics
Discrete feature assignment
Handling of outliers and missing values
Descriptive statistics of variables
Classification imbalance handling

Fig. 4. Preprocessing content.

In Eq. (4), R represents the recall rate. Eq. (5) is the expression for F1.

$$F1 = \frac{2P*R}{P+R} \tag{5}$$

Eq. (6) is the calculation of ROC.

$$FPR = \frac{FP}{FP+TN} \tag{6}$$

In Eq. (6), FPR represents the ROC value. To further explain the output results of the system, this study will use Shapley addition to explain the output results of the prediction system. As a visualization tool, it calculates each feature and explains the system's prediction results by contributing to the prediction. Eq. (7) is the calculation of the Shapley value.

$$y_i = y_{base} + f\left(x_{i1}\right) + f\left(x_{i2}\right) + \ldots + f\left(x_{ik}\right) \tag{7}$$

In Eq. (7), x_{ik} represents the k -the feature of the i -the sample. The system baseline is y_{base}. The predicted value of the system for this sample is y_i. $f\left(x_{ik}\right)$ represents the Shapley value of x_{ik} [23]. Generally speaking, $f\left(x_{ik}\right)$ is the contribution of x_{ik} to the final predicted value y_i. When $f\left(x_{ik}\right) > 0$, it indicates that the feature can improve the predicted value and has a positive effect. On the contrary, it indicates that the feature value leads to a decrease in the predicted value, which has a reverse effect. This value can reflect the influence and positive and negative shapes of sample features and has a powerful data visualization function, which is conducive to visually displaying the prediction results.

C. Recurrence Risk Classification of COPD Patients Based on K-means

The last layer of the model is the classification of recurrence risk for COPD patients. After selecting the best recurrence and readmission prediction method XGBoost, the selected method can be arranged and used. And combined with the predicted recurrence and readmission results of the patient, the discharge decision was obtained together. On this basis, a risk scoring system is constructed to divide recurrent COPD patients into different risk groups based on the obtained risk scores, providing decision support for doctors. The probability of XGBoost output on the test dataset was stratified using k-means and the results were validated in Fig. 5.

According to Fig. 5, the risk stratification process requires first using a complete prediction dataset to perform predictions on XGBoost. Then, the prediction probabilities output by XGBoost on the training dataset are grouped according to the equidistant method, and samples that do not meet the required probabilities are discarded. Then, the retained samples are input into XGBoost, and the prediction probability is clustered according to k-means. Samples that do not meet the required probability are eliminated, and samples that meet the probability continue to be predicted in XGBoost [24-25]. Finally, the evaluation index values obtained from each method training are calculated, and patients are stratified according to risk. In the k-means clustering, a dataset with d dimensions, whose number is n, is divided into k clusters. The purpose is to minimize the sum of squares of the distances between each observation in the same cluster and the center of the cluster to which it belongs in Eq. (8).

$$J\left(c,u\right) = \sum_{i=1}^{M} \left\| x_i - u_{ci} \right\|^2 \tag{8}$$

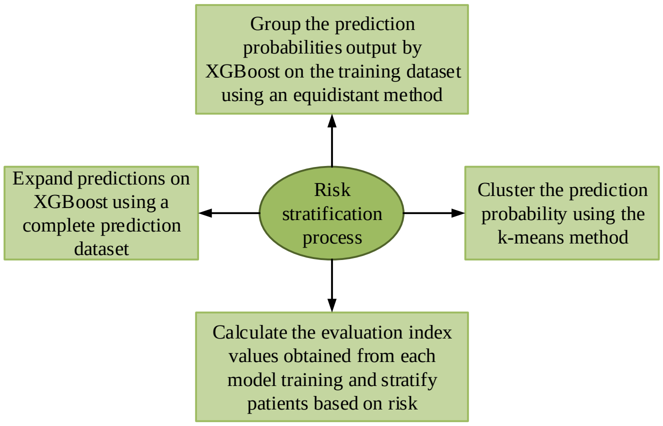

Group the prediction probabilities output by XGBoost on the training dataset using an equidistant method
Expand predictions on XGBoost using a complete prediction dataset
Risk stratification process
Cluster the prediction probability using the k-means method
Calculate the evaluation index values obtained from each model training and stratify patients based on risk

Fig. 5. Risk stratification process.

In Eq. (8), x_i represents the i -the sample. c_i is the cluster to which x_i belongs. u_{ci} represents the center point of the cluster pair. M is the total number of samples [26]. By using the k in the initial conditions, multiple different choices are made for k class, and the optimal clustering is obtained.

D. Methodology

In this study, the design of data collection and analysis programs is crucial. Data are collected from hospitals or medical databases, including electronic health records and clinical trial data. Information is collected from various aspects such as medical history, lifestyle, treatment response, and laboratory test results. The data types include patient basic information, clinical data, lifestyle data, and treatment information. Patients' basic information features include age, gender, weight, etc. Clinical data features include lung function test results, blood gas analysis, imaging examination results, etc. Treatment information features include drug treatment, non-drug treatment, etc. Lifestyle data features include smoking history, dietary habits, activity levels, etc.

Then the data are preprocessed and cleaned, including handling missing values and outliers. Standardization processing, such as normalizing numerical data, is conducted. Feature engineering, such as selecting features with high relevance, is performed. Statistical analysis of basic characteristics, such as age distribution and gender ratio, is carried out. The recurrence of COPD is analyzed, including frequency, severity, etc.

The first step of analyzing the program is to select a suitable ML model based on the characteristics of the data. The second step is to train and validate the model, dividing the dataset into training and testing sets. The model is trained on the training set and its performance is verified on the testing set. The model parameters are optimized using methods such as cross validation. The third step is to conduct feature importance analysis, analyzing the impact of each feature on predicting COPD recurrence and determining the most influential risk factors. The fourth step is to carry out risk classification, classify patients according to the predicted results, and set clinical decision guidelines corresponding to different risk levels. These steps need to be carried out in compliance with relevant data protection regulations and ethical standards.

IV. Result Analysis of Recurrence Prediction and Risk Classification Model for COPD Patients

The results of the ML-based recurrence prediction and risk classification model for COPD patients were analyzed, thereby promoting the improvement of medical structure and the utilization of medical resources. These results include the NER results of COPD patient electronic medical records based on optimized BiLSTM, the recurrence prediction results of COPD patients based on XGBoost, and the recurrence risk classification of COPD patients based on k-means.

A. NER Results of Electronic Medical Records for COPD Patients Based on Optimized BiLSTM

A suitable programming language was selected to implement the BiLSTM-CRF system and conduct the experiment. Appropriate development tools and libraries were installed to support system development. Electronic medical record data were preprocessed, including word segmentation, feature extraction, and label format conversion. It was ensured that the data format met the model input requirements and useful features were extracted for system training. When setting up the experimental environment, it ensures that all development tools, libraries, and data can be installed and configured correctly, enabling it to run smoothly. The BiLSTM-CRF system is built, and Table I shows the experimental environment settings.

Statistical analysis was conducted on the results of three recognition methods, BiLSTM, CRF, and BiLSTM-CRF, on the test set. All data are the average obtained from a 10-fold cross test. The overall experimental results were calculated in Table II.

According to Table II among these three methods, BiLSTM-CRF has the highest accuracy, regression rate, and F1 value, followed by BiLSTM and CRF. Overall, the accuracy rates of the three recognition methods BiLSTM, CRF, and BiLSTM-CRF are 84.36%, 83.67%, and 87.40%, respectively. Among these three methods, BiLSTM-CRF has the best performance. Subsequently, the recognition performance of different types of entities using the BiLSTM-CRF method was analyzed in Fig. 6.

According to Fig. 6, there is no significant difference in P and R values between these six entities naming recognition results. BiLSTM-CRF has the best recognition effect for both "comorbidities" and "examination indicators", with F1 greater than 90%. For F1, "lifestyle habit" is 74.00% and "course of disease" is 74.39%, indicating the difficulty in distinguishing between these two types. Based on the analysis of experimental results and annotated corpora, the effectiveness of entity recognition is closely related to the number of annotations and the degree of differentiation between different types of entities.

TABLE I. EXPERIMENTAL ENVIRONMENT

Number	Name	Type
(1)	Operating system	Windows 10
(2)	Development language	Python
(3)	CPU	TeslaP40
(4)	GPU	Intel Core i5-4210M @2.50GHz
(5)	Learning framework	Kera’s
(6)	Development platform	Jupiter Notebook 5.78

TABLE II. PERFORMANCE OF THREE MODELS

Entity Category	BiLSTM			CRF			BiLSTM-CRF		
	P (%)	R (%)	F1(%)	P (%)	R (%)	F1(%)	P (%)	R (%)	F1(%)
Complication	90.25	87.43	87.24	87.75	85.06	86.37	92.09	91.84	91.97
Inspection indicators	87.68	88.36	87.76	85.81	93.86	84.31	91.35	92.09	90.22
Habits and customs	70.35	60.19	62.73	66.42	41.79	50.89	74.83	75.47	74.00
Course of disease	62.76	53.86	62.77	52.66	36.06	41.39	72.21	73.78	74.39
Indicator value	80.39	78.64	68.37	77.47	65.75	53.89	84.14	84.17	88.82
Previous constitution	81.27	78.26	72.43	79.74	73.16	77.79	83.05	83.26	84.63
Whole	84.36	81.58	79.07	83.67	74.31	79.07	87.40	88.28	85.83

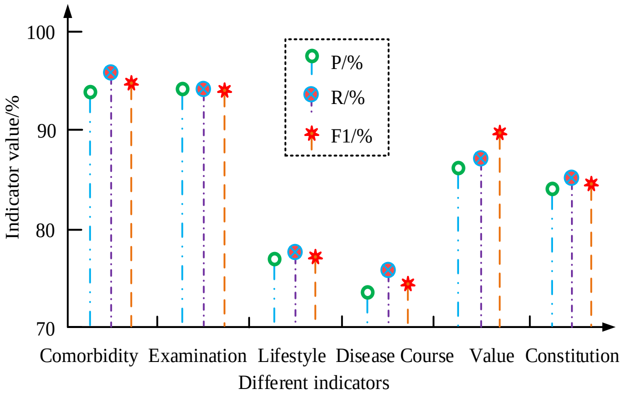


Fig. 6. The recognition effect of different types of entities in BiLSTM-CRF method.

B. *Prediction of Recurrence in COPD Patients Based on XGBoost*

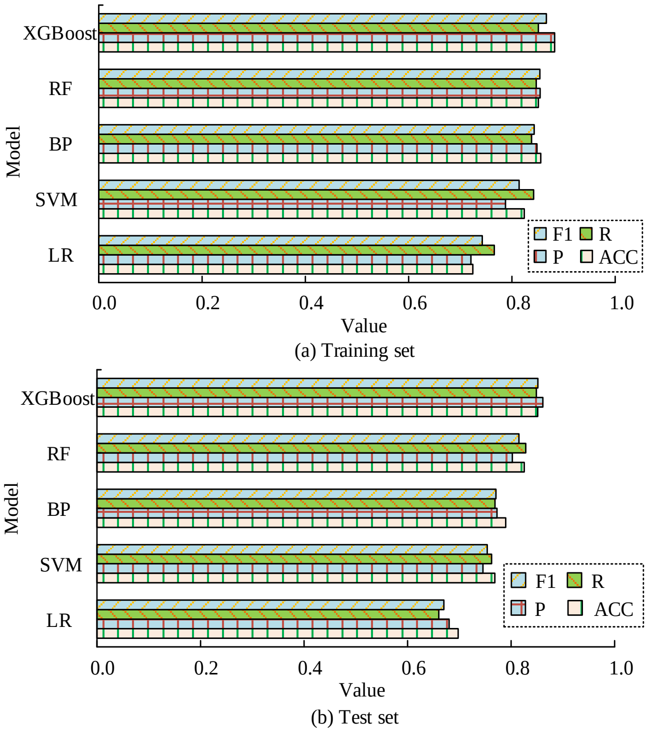
The experimental data are divided into training and testing sets, with a ratio of 7:3. The performance of individual patterns in the training set was evaluated using a grid search method with 10-fold cross validation. On this basis, the optimal hyperparameter combination for each method was selected, corresponding hyperparameter combinations were established, and fitting was conducted on a complete training dataset. Finally, the method was validated through independent experimental data. Due to the fact that this method is not affected by the experimental set during parameter adjustment and learning, the experimental set is only used for final evaluation. A comparison was made between five common ML constructed prediction systems: Logistic Regression (LR), Support Vector Machine (SVM), BP neural network, Random Forest (RF), and XGBoost in Fig. 7.

Fig. 7(a) and Fig. 7(b) show the predictive performance results of the five systems on the training and test sets, respectively. According to Fig. 7, five systems' accuracy on the training set is XGBoost, BP, RF, SVM, and LR, in descending order, with values of 0.8827, 0.8559, 0.8512, 0.8239, and 0.7244, respectively. The accuracy of these five systems on the test set is in descending order of XGBoost, RF, BP, SVM, and LR, with values of 0.8514, 0.8254, 0.7896, 0.7687, and 0.6978, respectively. Among these five systems, XGBoost has the best performance in terms of prediction accuracy, recall, and F1 value. Subsequently, the ROCs of five systems were analyzed in Fig. 8.

According to Fig. 8, the areas under the ROC of these five systems are XGBoost, RF, BP, SVM, and LR in descending order, with AUC values of 0.9173, 0.8256, 0.8252, 0.7682, and 0.6993, respectively. XGBoost has the best predictive ability and effectiveness, verifying its efficiency as a classification method for predicting the recurrence and readmission of COPD patients.

C. *Classification of Recurrence Risk in COPD Patients Based on K-means*

On the basis of XGBoost-based patient relapse readmission prediction system, k-means was selected for risk classification. To verify the superiority of this method, equidistant partitioning was used for comparative analysis. Firstly, patients are divided into five risk groups, namely low risk high reliability, low risk medium reliability, low risk or high-risk low reliability, high risk medium reliability, and high-risk high reliability. Among them, for patients with a lower risk of relapse and higher credibility, the classification probability tends to be closer to zero. For patients with high reliability and high risk of relapse and readmission, the classification probability tends to be closer to 1. Table III shows the results.



(a) Training set

(b) Test set

Fig. 7. Performance results of five common machine learning prediction system.

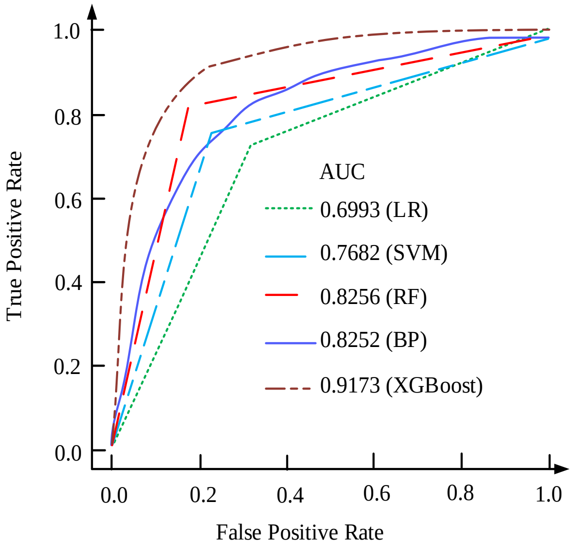


Fig. 8. ROC curves of 5 systems.

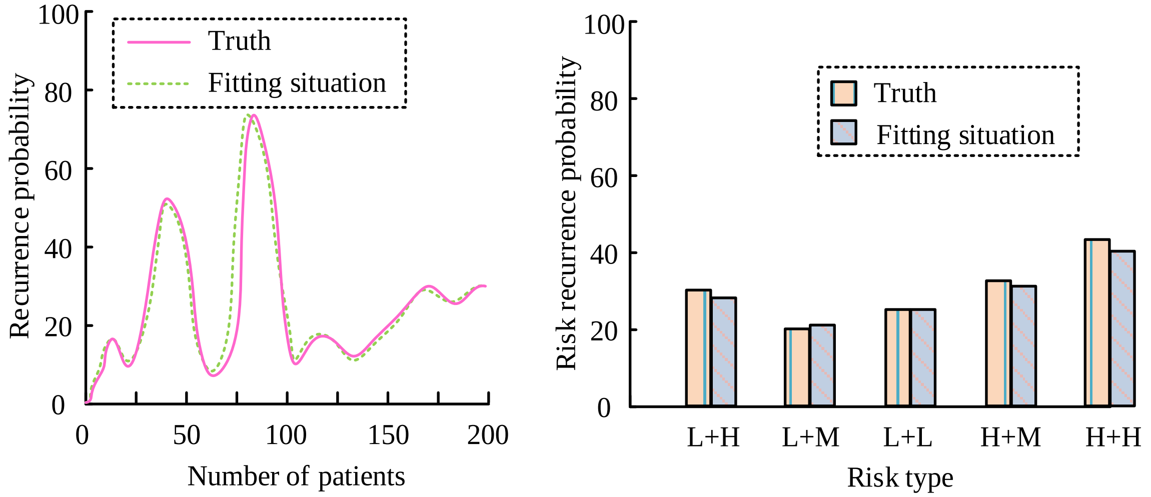
According to Table III, in equidistant stratification, only 17% of patients were classified into the low confidence group, which meant that the workload of manual evaluation was reduced by 83%. Among patients stratified using k-means, only 9% were classified into the low confidence group, which meant that the workload of manual evaluation was reduced by 91%. Compared with equidistant layering methods, k-means stratification required less manual evaluation workload. Therefore, k-means was selected as an evaluation indicator for the risk of relapse and readmission. In the current situation, if the balance point between different risk types is determined as a low risk or high-risk patient layer with a prediction probability of [0.286, 0.507] and low reliability, decision-makers only need to spend their limited resources to study these cases. That is, only 9% of the total cases need to be further manually evaluated. Through this approach, high-risk patients can achieve a balance between risk, cost, and resources.

Experimental validation confirms that among the three layers of the overall model for predicting recurrence and risk classification in COPD patients, each layer has the best performance. These three layers were integrated, and simulation analysis was conducted on the recurrence prediction and risk classification of COPD patients to verify the overall model's prediction and classification results in Fig. 9.

Fig. 9(a) and Fig. 9(b) show the fitting of the overall model for patient recurrence prediction and risk classification, respectively. According to Fig. 9, the simulation accuracy of the overall model for predicting patient recurrence and risk classification is 97.3% and 96.4%, respectively. This model can accurately predict the recurrence probability and risk classification of COPD patients, which is beneficial for providing decision-making assistance suggestions for hospitals.

TABLE III. PREDICTED RESULTS OF RISK GROUP FOR RECURRENT READMISSION PATIENTS

Method	Confidence level	Probability	Discard data (%)	Retained data	Accuracy (%)	Sensitivity (%)	Specificity (%)	Class I error	Class II error
Equidistant partition	All	All	0	1335	85.09	71.39	90.03	82	121
	Low	<0.4,>0.6	4	1287	83.92	71.82	91.63	70	115
	Medium	<0.3,>0.7	16	1110	86.48	73.24	94.75	46	103
	High	<0.2,>0.8	37	830	89.05	73.13	74.76	23	68
	Highest	<0.1,>0.9	76	332	94.32	76.52	99.62	2	19
K-means clustering and partitioning	All	All	0	1345	85.26	71.53	90.35	82	117
	Medium	<0.286,>0.507	10	1220	87.53	77.64	87.96	91	62
	Highest	<0.156,>0.759	46	751	90.21	82.94	93.56	33	42



(a) Recurrence risk prediction

(b) Risk type prediction

Fig. 9. Simulation effect of the overall model.

V. Discussion

Through the analysis of the results, the BiLSTM-CRF model performs the best in entity recognition tasks, followed by BiLSTM, and CRF performs the weakest. The BiLSTM-CRF model combines BiLSTM with CRF, the former can capture long-range dependent features of data, and the latter can consider constraints between labels in sequence prediction problems. This combination enables the model to perform well in entity recognition tasks with strong dependencies. Although BiLSTM performs well in capturing the contextual information of data, the lack of CRF layers makes it unable to utilize global information in label sequence prediction, resulting in weaker performance than the BiLSTM-CRF model. The simple CRF model only considers the relationship between labels and ignores contextual information, so its performance is the weakest. Entity recognition shows good recognition performance for "comorbidities" and "examination indicators", which may be due to the standardized expression of these entity types in the text. Entities such as "lifestyle habits" and "disease course" have poor recognition performance due to their diverse or vague expressions. The effectiveness of entity recognition is closely related to the quantity and quality of annotations, and more annotated samples and high-quality annotations can improve the accuracy of entity recognition.

The XGBoost model performs the best in predicting recurrence in COPD patients. XGBoost is an efficient gradient boosting algorithm that enhances model performance by continuously reducing errors during iterations. The high or low AUC value also reveals XGBoost's classification ability. In terms of risk classification, the k-means clustering algorithm reduces the workload of manual evaluation compared to equidistant partitioning, indicating that k-means can more effectively classify patients into different risk groups. This method supports hospitals to allocate resources more accurately, focusing on patients with moderate recurrence risk and lower credibility. Finally, the accuracy of the overall three-layer model in predicting recurrence and risk classification of COPD patients is verified through simulation analysis. This indicates that combining different analytical methods can comprehensively evaluate and predict the recurrence risk of patients, thereby providing more accurate support for hospital decision-making.

The entity recognition results indicate that the quantity and quality of annotated samples can significantly affect the accuracy of recognition. In the recurrence prediction model, XGBoost performs excellently with its efficient iterative performance. The selection of risk classification methods has a significant impact on reducing the workload of manual assessment, and the k-means method has shown advantages. Combining different methods can improve the accuracy of predictions and help hospitals make more precise decisions in resource allocation and decision support.

VI. Conclusion

Predicting the recurrence and hospitalization of frequent COPD patients that can be prevented, as well as understanding the reasons for recurrence and hospitalization, are important tasks to ensure the health of the people. This study uses ML to predict the recurrence probability of COPD patients and classify their risks, providing a scientific basis for medical institutions. These studies have confirmed that the accuracy rates of these three recognition methods BiLSTM, CRF, and BiLSTM-CRF are 84.36%, 83.67%, and 87.40%, respectively. BiLSTM-CRF has the best recognition effect for "comorbidities" and "examination indicators", with F1 greater than 90%. For F1, the "lifestyle habit" is 74.00% and the "course of disease" is 74.39%, indicating the difficulty in distinguishing between these two types. In the prediction of recurrence, the accuracy rates of XGBoost, BP, RF, SVM, and LR on the training set are 0.8827, 0.8559, 0.8512, 0.8239, and 0.7244, respectively. Their accuracy rates on the test set are 0.8514, 0.8254, 0.7896, 0.7687, and 0.6978, respectively. Among patients stratified using k-means, only 9% are classified into the low confidence group, which means that the workload of manual evaluation is reduced by 91%. Compared with the equidistant stratification method, the k-means stratification requires less manual evaluation workload, and the overall model has a more accurate fit for predicting recurrence and risk classification of COPD patients. There are still many shortcomings in this study, such as insufficient data scale and lack of data diversity. Future work will expand the dataset size to include more data from COPD patients to improve the model's generalization ability. And the diversity of data should be increased to ensure that patient data cover different geographical regions, races, genders, and age groups to enhance the comprehensiveness and adaptability of the model.

References

[1] Wang P X, Xu Y, Sun Y F, Cheng JW, Zhou K Q, Wu SY, Hu B, Zhang ZF, Guo W, Cao Y. Detection of circulating tumor cells enables early recurrence prediction in hepatocellular carcinoma patients undergoing liver transplantation. Liv. Int, 41(3):562-573(2021).

[2] Ye Z, Zhang Y, Liang Y, Lang J, Yang Cervical Cancer Metastasis and Recurrence Risk Prediction Based on Deep Convolutional Neural Network. Cur. Bio, 2022, 17(2):164-173.

[3] Chen Z. Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm. Jou. Com. Cog. Eng, 1(3): 103-108(2022).

[4] Kawahara D, Nishibuchi I, Kawamura M, Yoshida T, Nagata Y. Radiomic Analysis for Pretreatment Prediction of Recurrence after Radiotherapy in Locally Advanced Cervical Cancer. International Journal of Radiation Oncology, Biology, Physics, 2021,111 (3S):E93-E93.

[5] Xiaoke Z, Yu H, Liang Z, Tao L, Zhang M .A prognostic nomogram for predicting risk of recurrence in laryngeal squamous cell carcinoma patients after tumor resection to assist decision making for postoperative adjuvant treatment. Journal of Surgical Oncology, 2019,120(4):698-706.

[6] Chan L, Sadahiro S, Suzuki T, Okada K, Miyakita H, Yamamoto S, Kajiwara H.Tissue-Infiltrating Lymphocytes as a Predictive Factor for Recurrence in Patients with Curatively Resected Colon Cancer: A Propensity Score Matching Analysis.. Oncology, 2020, 98(10):680-688.

[7] Lafaie L, Thomas Célarier, Goethals L, Pozzetto B, Botelho ㎞ Evers E. Recurrence or Relapse of COVID-19 in Older Patients: A Description of Three Cases. Journal of the American Geriatrics Society, 2020,68(10):2179-2183.

[8] Dowsett M. Integration of Clinical Variables for the Prediction of Late Distant Recurrence in Patients with Estrogen Receptor- Positive Breast Cancer Treated With 5 Years of Endocrine Therapy: CTS5 (vol 14, pg 234, 2019). Journal of Clinical Oncology, 2020,38(6):656-656.

[9] Postiche H, Piccard M. BiLSTM-SSVM: Training the BiLSTM with a Structured Hinge Loss for Named- Entity Recognition. IEEE transactions on big data, 8(1):203-212(2022).

[10] Benali B A, Mihi S, Moku A, Bazi I EI, Yakhouba N. Arabic named entity recognition in social media based on BiLSTM-CRF using an attention mechanism. Journal of Intelligent & Fuzzy Systems: App. Eng. Tec, 42(6):5427-5436(2022).

[11] Long R, Yang D, Liu Y. Disease Net: A Novel Disease Diagnosis Deep Framework via Fusing Medical Record Summarization. IAE. Int. Jou. Com. Sci, 49(3 Pt.2):808-817(2022).

[12] Puh K, Babac M B. Predicting sentiment and rating of tourist reviews using machine learning. Jou. Hos. Tou. Ins, 6(3):1188-1204(2023).

[13] Matheson A M, Parraga G. Machine Learning Predictions of COPD Mortality: Com.Hid. See. Che, 158(3):846-847(2020).

[14] AL khadar H, Meluskey M, White S, Ellis I, Gardner A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. Jou. Ora. Pat& Med, 50(4):378-384(2021).

[15] Wong N C, Lam C, Patterson L, Shay Egan B. Use of machine learning to predict early biochemical recurrence after robot-assisted prostatectomy.BJU International, 123(1):51-57(2019).

[16] Paredes A Z, Hyer J M, Salimgarh D I, Moro A, Pawlik TM. A Novel Machine-Learning Approach to Predict Recurrence After Resection of Colorectal Liver Metastases. Ann. Sur. Onc, 27 (13):5139-5147(2020).

[17] Zhang S, Zhu H, Xu H, Zhu G, Li K C. A named entity recognition method towards product reviews based on BiLSTM-attention-CRF. Int. Jou. Com. Sci. Eng, 2022,25(5):479- 489.

[18] Li D, Dong C, Chen Z, Dong Y, Liu J. A combinatorial machine-learning-driven approach for predicting glass transition temperature based on numerous molecular descriptors. Mol. Sim, 49 (6):617-627(2023).

[19] Guo Y, Mustafa Z, & Kaunda D. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. Jou. Com. Cog. Eng, 2(1), 5–9(2022).

[20] Liu M, Stella F, Homeroom A, Lucas, Peter J F, Lonneke B, Bischoff E.A comparison between discrete and continuous time Bayesian networks in learning from clinical time series data with irregularity. Art. Int. Med, 95(APR.):104-117(2019).

[21] Reito A, Karola K, Pekkanen L, Palomera J. 30-day recurrence, readmission rate, and clinical outcome after emergency lumbar discectomy. Spine, 45(18): 1253-1259(2020).

[22] Lao Y, Yu V, Pham A, Wang T, Sheng K. Quantitative Characterization of Tumor Proximity to Stem Cell Niches: Implications on Recurrence and Survival in GBM Patients. Int. Jou. Rad. Ons. Bio. Phys, 110(4):1180-1188(2021).

[23] Meng T, Huang R, Hu P, Yin H, Song D. Novel Nomograms as Aids for Predicting Recurrence and Survival in Chordoma Patients: A Retrospective Multicenter Study in mainland China. Spine, 46(1): E37-E47(2020).

[24] Lei M, Han Z, Wang S, Han T, Fang S, Lin F, Huang T. A machine learning-based prediction model for in-hospital mortality among critically ill patients with hip fracture: An internal and external validated study. Injury, 54(2):636-644(2023).

[25] Holtkamp L H J, Lo S N, Thompson J F, Spillane A J, Stretch J R, Saw R P M, Shannon K F, Newegg O E, Hong A M. Adjuvant radiotherapy after salvage surgery for melanoma recurrence in a node field following a previous lymph node dissection. Jou. Sur. Ons, 128(1):97-104(2023).

[26] FangY, LuoB, Zhao T, He D, Jiang B, Liu Q. ST-SIGMA: Spatial-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting. CAAI Transactions on Intelligence Technology, 7(4):744-757(2022).

Supply Chain Disturbance Management Scheduling Model Based on HPSO Algorithm

Ling Wang*
Nanjing University of Finance & Economics Hongshan College, Nanjing, 210003, China

Abstract*—The continuous expansion of business has led to the development of enterprises from vertical integration to horizontal integration, and the interlocking of the supply chain system, but the influence of anti-production behavior factors and the frequent occurrence of disruption events lead to difficulties in supply chain scheduling, which affects the development of enterprises. To address the above problems, the study analyzes the factors influencing counterproductive behavior based on system dynamics, constructs a supply chain disruption management scheduling model on this basis, and solves the supply chain disruption management scheduling model using Hybrid Particle Swarm Optimization algorithm. The findings indicate that the number of non-inferior solutions, uniformity of distribution of non-inferior solutions, dominance ratio of non-inferior solutions, average distance between non-inferior solutions and optimal Pareto, maximum distance, dispersion of non-inferior solutions and coverage of non-inferior solutions of the hybrid particle swarm algorithm are 12.3, 5.283, 0.264, 0.611, 4.474, 4.627, 601.300, respectively in the A condition, 601.300. The number of non-inferior solutions, uniformity of non-inferior solution distribution, dominance ratio of non-inferior solutions, average distance between non-inferior solutions and optimal Pareto, maximum distance, dispersion of non-inferior solutions and coverage of non-inferior solutions for the hybrid particle swarm algorithm under B condition are 12.3, 5.283, 0.264, 0.611, 4.474, In summary, the proposed algorithm has excellent performance and can effectively reduce the impact of interference events, thereby improving the level of supply chain interference management and scheduling, and promoting the sustainable development of this field.

Keywords*—*HPSO algorithm; disturbance management; supply chain; system dynamics; anti-production behavior

I. Introduction

With the popularization of computer technology and the development of market economy, enterprises are shifting their focus from themselves to supply chain, and supply chain management is particularly important for enterprise management [1-2]. Supply chain management is a management mode that effectively organizes suppliers, manufacturers and distributors to jointly complete product production, transportation and distribution on the premise of minimizing the cost of the entire supply chain [3-4]. However, the supply chain system itself is a complex and dynamic network, which is faced with many challenges from uncontrollable factors both internally and externally, such as resource shortage, equipment failure, and market demand fluctuations. These factors increase the uncertainty and dynamics of the supply chain, making it difficult for traditional supply chain management methods to effectively cope with [5]. At present, how to accurately identify and quantify the influencing factors of employee Counterproductive Work Behavior (CWB) in the supply chain environment and its impact on supply chain performance is a relatively difficult problem [6-7]. In order to better deal with the deterioration effect of CWB and the scheduling deviation of SCM disturbance, in this study, the influence factors of CWB are deeply analyzed, and the corresponding mathematical model is established by using the theory of System Dynamics (SD) to quantify the impact of CWB on supply chain performance. In addition, the research also aims to solve the hybrid model using HPSO algorithm to verify the validity and practicability of the model and provide decision support for enterprises in the face of supply chain disturbance events. Finally, through empirical research, the effect of the proposed model and algorithm in practical application is verified, which provides theoretical basis and practical guidance for the supply chain management of enterprises. This study will integrate SD theory and HPSO algorithm, enrich the research methods and technical means in the field of supply chain management, and provide new perspectives and ideas for the development of supply chain management theory. In addition, this study will provide decision support and practical guidance for enterprises to deal with supply chain disturbance events in actual operations. By optimizing the supply chain scheduling strategy, enterprises will be able to improve the stability and efficiency of the supply chain, reduce operating costs, improve customer satisfaction, and thus enhance the market competitiveness of enterprises. Finally, the research on employee CWB will help enterprises to understand the impact of employee behavior on supply chain performance more comprehensively, and then take targeted management measures to reduce potential losses.

This research is mainly divided into four sections, Section II is a review of relevant research results, Section III is the use of HPSO algorithm to solve the SCIMC model, Section IV is to verify the effectiveness of the HPSO algorithm proposed by the research, and Section V and Section VI is discussion and summary of the research respectively.

II. Related Work

The traditional SC optimization research is mostly focused on transportation and distribution, but rarely involves production optimization algorithm. The findings indicate that the algorithm can effectively reduce the time for computation [8]. Li et al. reduce the maximum time for completion by establishing a task pool and employing a genetic forbidden search algorithm for the production and transportation integration scheduling problem in a hybrid flow shop.

Experimental results confirm that the method can successfully address the scheduling problem for production and transportation integration [9]. Goli et al. employ a cycle duration and integer multiplier technique to coordinate the replenishment cycle of the SC and developed a simulated annealing algorithm to solve the economic lot size and delivery scheduling problems of a multi-stage SC. Simulation experiments demonstrated that the method can reduce the production cost of the SC [10]. Solina et al. proposed a quantitative approach to production and distribution to minimize production and distribution costs with reference to real-life food companies. The findings indicate that the method can significantly improve the performance and sustainability of the SC [11]. Du et al. explored the importance of dynamic optimal production scheduling with demand fluctuations and uncertainties. The study developed a multi-objective genetic algorithm for precast manufacturing based on a dynamic flow shop scheduling model, and simulation studies demonstrated that the method can cope effectively with demand changes [12].

Disruption management is also needed to solve the stochastic disturbance problem because of the real-life uncontrollable factors and SC systems are often affected by many disruptive events that delay production schedules and increase production costs. Lee J et al. designed a joint reactive and proactive airline disruption management method to cope with air traffic disruptions that lead to flight delays, cancellations, and missed passenger connections. The method can predict the probability of future disruptions by estimating the system delays at hub airports. The research data showed that the method can effectively reduce the expected recovery cost of airlines [13]. Jiang et al. constructed an attitude based disruption management model for handling delivery delays to minimize the negative impact of sudden disruptions in the distribution phase of the SC, and designed a heuristic algorithm, and the simulation results verified the effectiveness of the method [14]. Ning et al. addressed the flexible job shop in a comparison experiment with the traditional rescheduling method, the interruption management method improves the stability of the production and processing system under the deterioration effect [15]. Pandi et al. developed a GPU-based adaptive large-neighborhood search technique to address the issue of fleet interruption due to vehicle failure. Simulation experiments indicate that the algorithm can reduce the idle time and operating cost of the fleet under normal operation [16]. Malik A I et al. present a production disruption model for a multi-product single-stage production inventory system to handle the problem of unforeseen disturbances disrupting the entire manufacturing schedule. The data suggest that the effectiveness and superiority of the performance of this production disruption model [17].

In summary, there are many research results about interference management in the field of SC production scheduling, but the majority of the research findings center on the reasonable distribution of negotiated benefits and the optimization of the overall benefits of the SC, ignoring the impact of interference events and CWB of employees, which leads to difficulties in SC scheduling. To address the problem that SC scheduling is easily affected by disruptive events and CWB, this paper analyzes the influencing factors of CWB based on SD theory and establishes SCIMC model, and solves the model by HPSO algorithm.

III. Construction of SCIMC Model based on HPSO Algorithm

The analysis of CWB influencing factors based on SD theory is the premise of SCIMC model construction. Since the SCIMC problem has multi-objective and non-linear characteristics, the study introduces the HPSO algorithm to solve it, and this chapter focuses on the analysis of CWB influence factors based on SD theory and the design of the HPSO algorithm.

A. Analysis of the Factors Influencing Employee CWB Based on SD Theory

SD theory is an effective tool that combines cybernetics, systems theory and information theory and uses computer simulation techniques to study the feedback structure and behavior of systems. It mainly studies the dynamic development law of the system through modeling, simulation and comprehensive reasoning according to the feedback characteristics that the internal components of the system are causal [18]. The main steps of SD theory are shown in Fig. 1.

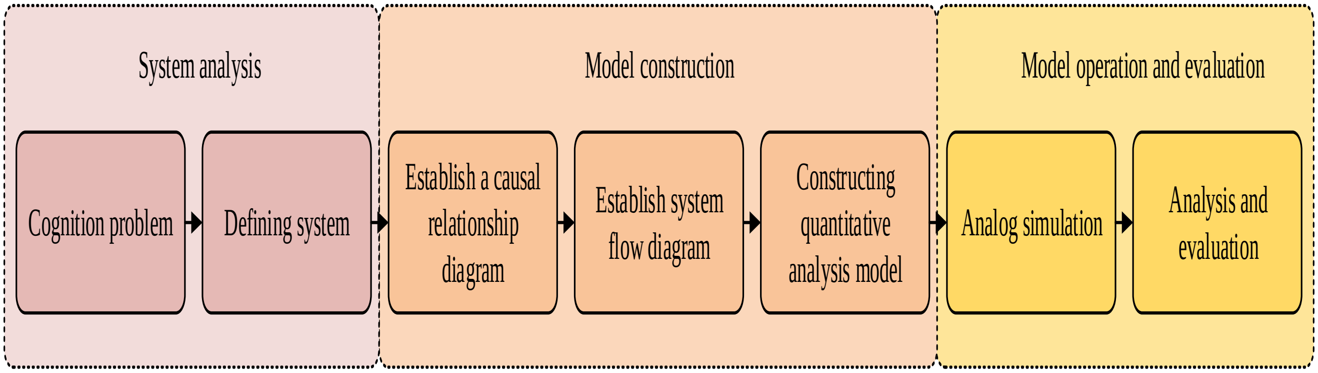

System analysis
Model construction
Model operation and evaluation
Cognition problem
Defining system
Establish a causal relationship diagram
Establish system flow diagram
Constructing quantitative analysis model
Analog simulation
Analysis and evaluation

Fig. 1. The main steps of SD theory.

The SD theory in Fig. 1 is mainly composed of three steps: system analysis, model construction, model operation and evaluation. The presence of CWB frequently disrupts the normal running of the production line [19]. To dynamically describe the nonlinear mechanism of CWB, the study uses the SD method as an entry point to explore the influencing factors of CWB. the influencing factors of CWB include job satisfaction A1, sense of organizational justice A2, supervision level A3, team atmosphere positivity A4, group normative level A5, organizational culture building A6, and organizational concern A7. The causal relationship of each influencing factor of CWB is shown in Fig. 2.

Fig. 2 depicts the CWB feedback relationship, which is separated into six major sections. In the first step, which functions as positive feedback, higher job satisfaction lowers the likelihood of CWB incidence, which causes the amount of group norms to rise, and a corresponding decrease in the level of job boredom as an important component of job satisfaction, which further increases job satisfaction. The second part is a negative feedback process, where an increase in organizational justice causes an increase in employee job satisfaction, which decreases the chance of CWB occurrence and leads to an increase in the level of employee burnout, which decreases organizational justice and concern. The third part is the negative feedback process, with the improvement of the organizational supervision mechanism, the team atmosphere positivity and the number of behavior correction is increasing, accompanied by a decrease in the sense of organizational justice and attention, which will bring negative impact on the level of organizational supervision. The fourth part is the positive feedback process, the higher the team climate positivity, the lower the level of group regulation will be, accompanied by employee alienation and increasing job conflict will also lead to a decrease in job satisfaction, which will increase the chances of CWB, further improving the level of organizational regulation and thus increasing the team climate positivity. The fifth part is the positive feedback process, where the increase in the level of group norms makes CWB less likely to occur, thus reducing the level of organizational attention and supervision, making the team gradually looser and further increasing the level of organizational norms. The sixth part is a negative feedback process. The improvement of organizational culture also increases the motivation of the team atmosphere, so the chance of CWB decreases, but the decrease of CWB makes the organization slack, which further reduces the level of attention and organizational culture. The study adds the corresponding state variables, rate variables and auxiliary variables based on the feedback relationship of CWB-related influencing factors, and draw the CWB-SD model flow diagram using Vensim simulation software.

$$\begin{cases} Z_{A1} = 2 + \int_{t0}^{t} \left(H_1 - H_2\right) dt \\ Z_{A2} = 1 + \int_{t0}^{t} \left(H_3 - H_4\right) dt \\ Z_{A3} = Z_{A2} \times B_1 + Z_{A6} \times B_2 + Z_{A8} \times B_3 + Z_{A9} \times B_4 \\ Z_{A4,5} = 1.5 + \int_{t0}^{t} \left(H_5 - H_6\right) dt \\ Z_{A6} = 2 + \int_{t0}^{t} H_9 dt \end{cases} \tag{1}$$

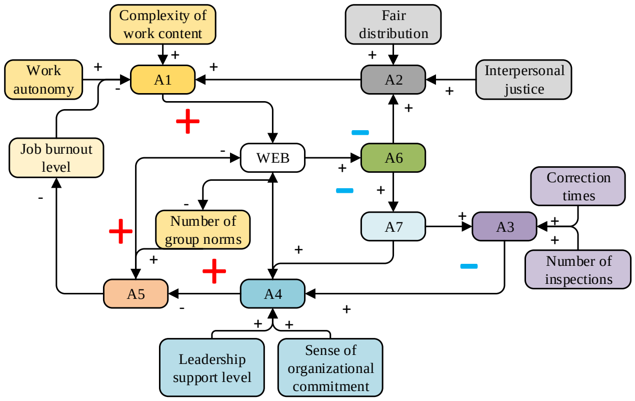


Fig. 2. Feedback relationship between CWB influencing factors.

In Eq. (1), Z_{A1} to Z_{A6} represent job satisfaction, organizational justice, organizational supervision level, team atmosphere motivation, group norm level, and organizational culture building level, respectively. B_i H_1 and H_2 represent rates of satisfaction growth and decline, H_3 and H_4 represent the fairness's growth and decline rates, and H_5 and H_6 represent the increase and decrease rates of organizational culture building. The expressions of the number of behavioral corrections Z_{A8} and the monitoring and improvement mechanism Z_{A9} are shown in Eq. (2).

$$\begin{cases} Z_{A8} = 2 + \int_{t0}^{t} H_{10} dt \\ Z_{A9} = 2 + \int_{t0}^{t} H_{11} dt \end{cases} \tag{2}$$

In Eq. (2), H_9 , H_{10} refer to the rate of increase in the level of organizational culture building and the rate of increase in behavior modification, respectively. The expression of CWB can be obtained from Eq. (1) and Eq. (2), see Eq. (3).

$$Z_{CWB} = Z_1 \times G_1 + Z_2 \times G_2 + Z_3 \times G_3 + Z_4 \times G_4 + Z_5 \times G_5 + Z_6 \times G_6 \tag{3}$$

In Eq. (3) G_i denotes the influence coefficient of each state variable. Human behavior influences the operation and state of the system, and employees' dissatisfaction with their jobs directly leads to an increased chance of CWB and thus negativity. To portray the effect of subjective human behavior on CWB, the study uses employee dissatisfaction to describe the deterioration rate and thus measure the processing time after disturbance, and the dissatisfaction value $Q\left(R_i\right)$ is expressed in the range of $x_i < 0$ as shown in Eq. (4) [20].

$$Q(R_i) = \chi (R_i - O_i)^{\beta} \tag{4}$$

In Eq. (4), O_i denotes the initial scheduling scheme and χ is the risk aversion factor. β It refers to the degree of concavity of the value curve, and its value range is $(-\infty, 1)$. When $Q(R_i) = 1$, $R_i = O_i + \left(\frac{1}{\chi}\right)^{\frac{1}{\beta}}$, the dissatisfaction function $Q(x_i)$ see Eq. (5).

$$Q(x_i) = \begin{cases} 1, x_i \geq R_i \\ \chi (x_i - O_i)^{\beta}, O_i \leq x_i \leq R_i, i = 1, 2, 3, \cdots, n \\ 0, 0 \leq x_i \leq O_i \end{cases} \tag{5}$$

The occurrence of a disturbance event causes O_i to be no longer optimal, and the repair scheduling solution derived according to the specified constraint affects the change in machining position of the corresponding workpiece. The size of the measured disturbance can be expressed in terms of the amount of machining position change, and the dissatisfaction function of the amount of position disturbance $Q(s_j)$ is shown in Eq. (6).

$$Q(s_j) = \begin{cases} 1, s_j \geq R_1 \\ \chi_1 s_j^{\beta_1}, 0 \leq s_j < R_1 \end{cases} \tag{6}$$

In Eq. (6) s_j denotes relative position perturbation, $R_1 = \left(\chi_1^{-1}\right)^{\beta_1^{-1}}$ denotes the upper limit of dissatisfaction tolerance, and employee dissatisfaction Q is shown in Eq. (7).

$$Q = \sum_{j=1}^{n} Q(s_j) \Big/ n \tag{7}$$

Employee psychological dissatisfaction brings about defiance, which lengthens the operation's processing time and subsequently affects the deterioration rate. The deterioration rate function $\theta(s_j)$ is shown in Eq. (8).

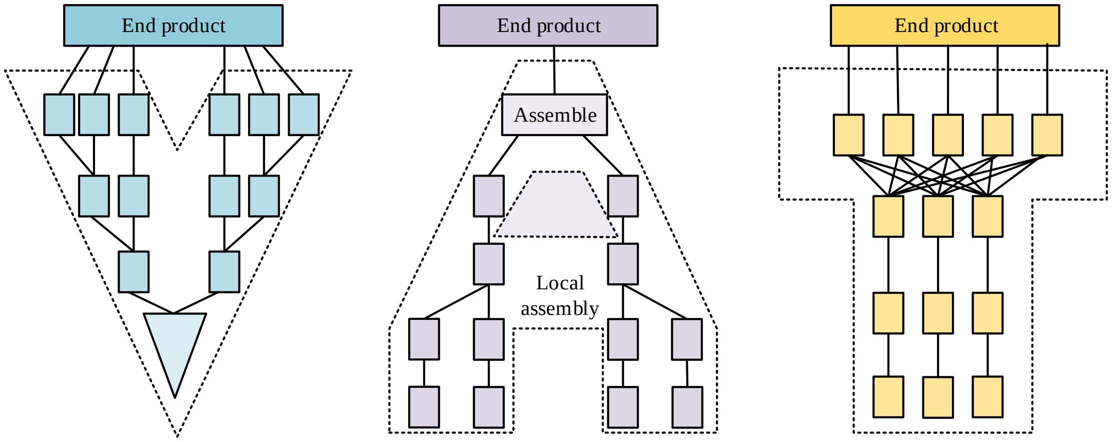
$$\theta(s_j) = \begin{cases} 1, s_j > R_1 \Box R_2 \cap s_j > R_1 \\ \chi_2 s_j^{\beta_2}, 0 \leq s_j < R_1 \Box R_2 \cap 0 \leq s_j < R_1 \end{cases} \tag{8}$$

In Eq. (8) $R_2 = \left(\chi_2^{-1}\right)^{\beta_2^{-1}}$, then the deterioration rate of operation time θ is shown in Eq. (9).

$$\theta = \sum_{j=1}^{n} \theta(s_j) \Big/ n \tag{9}$$

B. SCIMC Model Construction Based on HPSO Algorithm

Each node enterprise in SC rotates around the core enterprise, forming a fully functional network chain, through regulating the flow of information and cash from acquiring raw resources to producing finished goods, and finally delivering products to consumers through the sales network. According to the different products and manufacturing processes, the SC is split into V-type, T-type and A-type, and the basic structure of the SC is shown in Fig. 3 [21].



(a) Basic structure of V-type supply chain

(b) Basic structure of a-type supply chain

(c) Basic structure of t-type supply chain

Fig. 3. Basic structure of SC.

Fig. 3(a) shows the basic structure of V-type SC, which is the most basic structure in the SC mesh. The success of V-type SC depends on the reasonable arrangement of the critical internal capacity bottlenecks. Fig. 4(b) depicts the basic structure of A-type SC, the overall form of this SC is expressed as convergence type, and A-type SC is generally driven by orders and customers. No market forecast is taken. Fig. 4(c) shows the basic structure of T-type SC, which is a hybrid SC structure that mainly determines the manufacturing standardization of common parts to reduce the complexity. In a two-stage SC with a single manufacturer and supplier, both need to share the task of a batch of orders from customers, the supplier processes the relevant spare parts according to customer demand, and the manufacturer produces according to the spare parts delivered by the supplier, in which the supplier is in a dominant position and the supplier should be satisfied by the manufacturer on each requirement of the order. To increase production effectiveness, the study constructs the SCIMC model, which mainly consists of three parts: initial scheduling, interference management, and cooperation gain, and the initial scheduling is shown in Eq. (10).

$$\min\left\{ f\left(\pi_0{}^s\right)=\sum\nolimits_{j=1}^{n} w_j \cdot C_j,\ f\left(\pi_0{}^m\right)=\sum\nolimits_{j=1}^{n} w'_j \cdot C'_j \right\} \tag{10}$$

Eq. (10) is the optimisation goal for initial scheduling, $\pi_0{}^s$ and $\pi_0{}^m$ refer to the initial scheduling time of suppliers and manufacturers, respectively, w_j and C_j refer to the weighting factor and completion time of suppliers, respectively, and w'_j and C' refer to the weighting factor and completion time of manufacturers, respectively. The expression of interference management is shown in Eq. (11).

$$\min\left\{ f_1\left(\pi'\right)=\sum\nolimits_{j=1}^{n} w'_j \cdot C'_j,\ f_2\left(\pi'\right)=\sum\nolimits_{j=1}^{n} w'_j \cdot \bar{\Delta} t'_0 \right\} \tag{11}$$

Eq. (11) is the optimization objective of disturbance management scheduling when the machine is disturbed. $f_1\left(\pi'\right)$ denotes the optimization objective of the manufacturer's balanced disturbance repair solution and initial scheduling solution, and $f_2\left(\pi'\right)$ denotes the minimization objective, where $\bar{\Delta} t'_0=\max\left\{C'_j-\bar{C}'_j,0\right\}$, $\bar{C}'_j$ are the completion times of the manufacturer's artifacts J_j in the initial scheduling solution. The expression of the cooperative gain is shown in Eq. (12).

$$\min\left\{ f_3\left(\pi'\right)=-V_m \cdot V_s \right\} \tag{12}$$

Eq. (12) is the revenue maximization objective of supplier-manufacturer cooperation, and V_m and V_s denote the revenue of the manufacturer and supplier after the perturbation, respectively. The supplier's processing artifacts arrive before the manufacturer can start production, and the expression is $D_j^s \le S_j^m$, where D_j^s denotes the supplier's delivery time and S_j^m is the manufacturer's processing start time. In a product's processing system, neither the supplier nor the manufacturer is allowed to start both workpieces at the same time, see $\left(S_j \ge C_k\right) \vee \left(S_k \ge C_j\right), \forall j,k \in J$. During the time window when the disturbance occurs, the supplier cannot schedule the disturbed workpiece for processing, as expressed in the formula at $S_j^s, C_j^s \notin \left[t_1,t_2\right], \left(j \in list\right)$, where S_j^s and C_j^s denote the supplier's processing time and completion time, respectively. To efficiently optimize the SCIMC model, the study selects the HPSO for solving. Based on the PSO algorithm's great global fast search capability, the HPSO combines the strengths of the variable domain search algorithm's strong local fine search capability, and incorporates the heuristic algorithm obtained from the variational crossover theory related to the genetic algorithm. the HPSO algorithm flow is shown in Fig. 4 [22].

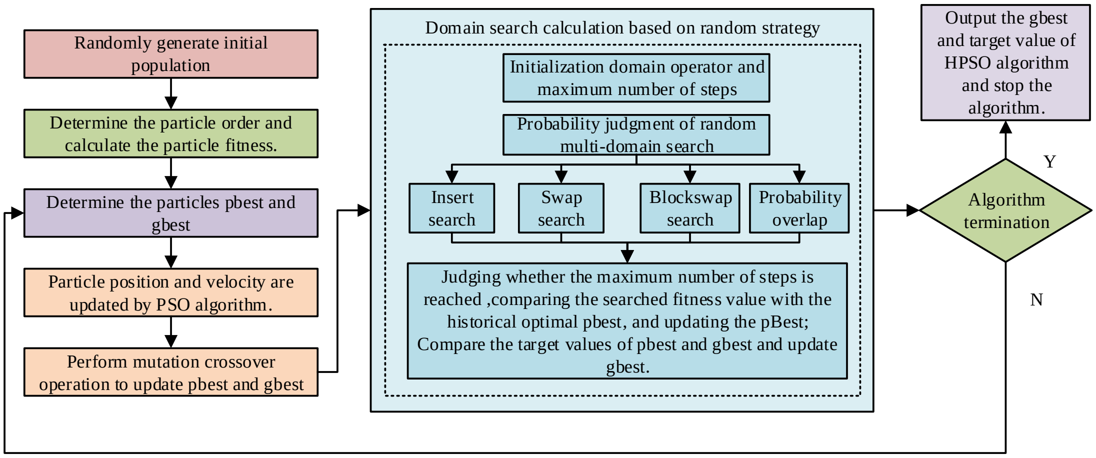

Randomly generate initial population
Determine the particle order and calculate the particle fitness.
Determine the particles pbest and gbest
Particle position and velocity are updated by PSO algorithm.
Perform mutation crossover operation to update pbest and gbest
Domain search calculation based on random strategy
Initialization domain operator and maximum number of steps
Probability judgment of random multi-domain search
Insert search
Swap search
Blockswap search
Probability overlap
Judging whether the maximum number of steps is reached ,comparing the searched fitness value with the historical optimal pbest, and updating the pBest; Compare the target values of pbest and gbest and update gbest.
Algorithm termination
Y
N
Output the gbest and target value of HPSO algorithm and stop the algorithm.

Fig. 4. HPSO algorithm flow.

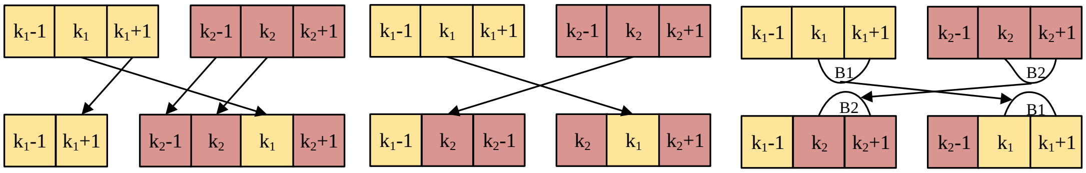
The HPSO algorithm in Fig. 4 consists of the basic PSO algorithm, variational operations, crossover operations and random field structures. The PSO algorithm first requires particle initialization and particle position initialization, where a particle represents a processing ordering and the particle initialization is to ensure that the processing of a workpiece is unique at the same time. Eq. (13) displays the iterative equation for a particle's velocity at the instant of $t+1$.

$$v_{i,j}(t+1)=wv_{i,j}(t)+c_1r_1\left[p_{i,j}-x_{i,j}(t)\right]+c_2r_2\left[p_{g,j}-x_{i,j}(t)\right] \tag{13}$$

In Eq. (13), w denotes the inertia factor, c_1,c_2 is the learning factor, r_1,r_2 is the arbitrary number generated between $(0,1)$, $p_{i,j}$ and $p_{g,j}$ denote the current and the global optimal position of the particle in the j dimension, respectively, so the iterative formula for the position of the particle at the moment of $t+1$ is Eq. (14)

$$x_{i,j}(t+1)=x_{i,j}(t)+\left|v_{i,j}(t+1)\right|, j\in 1,2,\cdots,n \tag{14}$$

After the initialization of the particle positions, the study also needs to form two new particles by mutating the particle's own best position $pbest$ and the population's best position $gbest$, and the mutated particles are used as the parents to perform the crossover operator based on the process encoding. The PSO algorithm after mutation crossover improves the capability of global search, but the capacity for local search still needs to be improved. To solve this problem, the study introduces a local search strategy with a random domain structure. The stochastic domain structure mainly consists of insertion domain, exchange domain and block exchange domain structure, and its structure is shown in Fig. 5.



(a) Insert domain structure

(b) Swap domain structure

(c) Blockswap domain structure

Fig. 5. Schematic diagram of domain structure.

Fig. 5(a) shows the insertion field structure, randomly inserting the artifact l before the artifact l_1 , where l_1 is any position in the arrangement π before l_1 . Fig. 6(b) shows the swap field structure, where the positions of the workpieces l_1 and in the arrangement l_2 π are randomly swapped. Fig. 6(c) shows the block swapping domain structure, randomly swapping the positions of the B_1 and B_2 blocks in the arrangement π . The local search strategy based on the random domain structure is to perform the domain operation with random probability c_{pm} and the random probability c_{pm} is shown in Eq. (15).

$$c_{pm}=\begin{cases}(\alpha_1\le r_1\le\beta_1)\Rightarrow insert\\(\alpha_2\le r_2\le\beta_2)\Rightarrow swap\\(\alpha_3\le r_3\le\beta_3)\Rightarrow blockswap\end{cases} \tag{15}$$

The probability interval overlap is defined in Eq. (15) as $COM<I,S,BS>$, then the priority levels of the domain operations are, in order, the insertion domain, the insertion domain and the block exchange domain.

IV. Performance Analysis of HPSO Algorithm in SCIMC Model

To evaluate how well the HPSO performs in the SCIMC model, the study first requires an analysis of the degree of variation of the WBB under different influencing factors, the experimental software is Visual and the time parameter is set to 20 in months.

Fig. 6 displays the CWB outcomes in various affecting circumstances. From the figure, the CWB changes with time and the degree of CWB change varies under different influencing factors. At the beginning of the simulation, CWB shows an increasing trend with time and reaches saturation at the 7th month, and then decreases rapidly after seven months, and the decreasing trend slows down at the 14th month. Job satisfaction has the greatest impact on CWB, and the level of organizational culture building has the least impact. To optimize the SCIMC model, the study conducted 20 independent numerical experiments on the HPSO algorithm, GA-TOM algorithm, basic PSO algorithm, and ACO algorithm, respectively. The HPSO algorithm's parameters were set with 100 iterations as follows, $w=0.2$, $c_1=c_2=2$, and $c_{pm}=\left\{[0,0.75],[0.55,0.95],[0.8,0.1]\right\}$

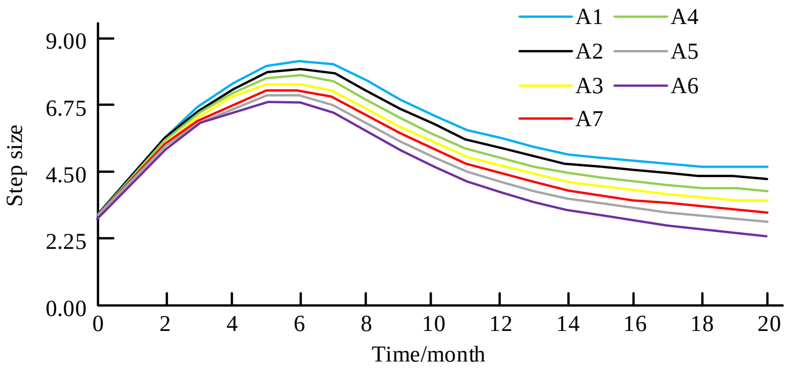


Fig. 6. Comparative analysis of CWB under different influencing factors.

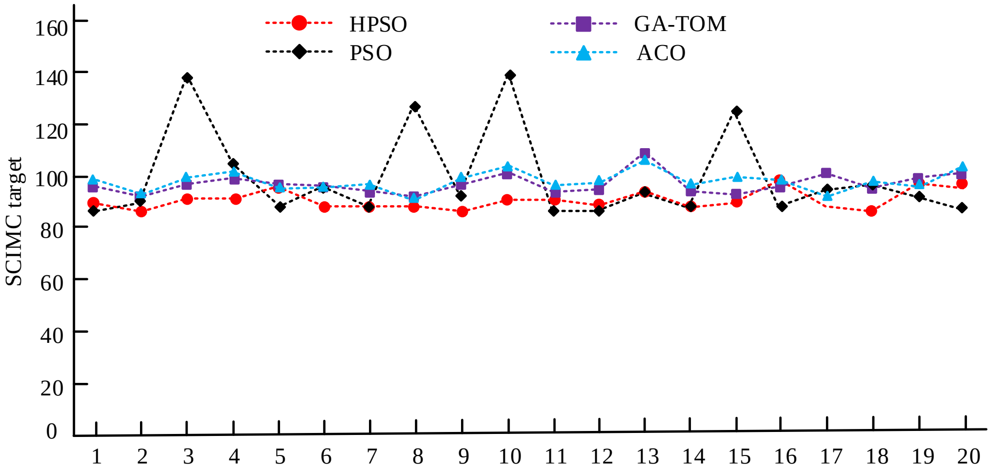


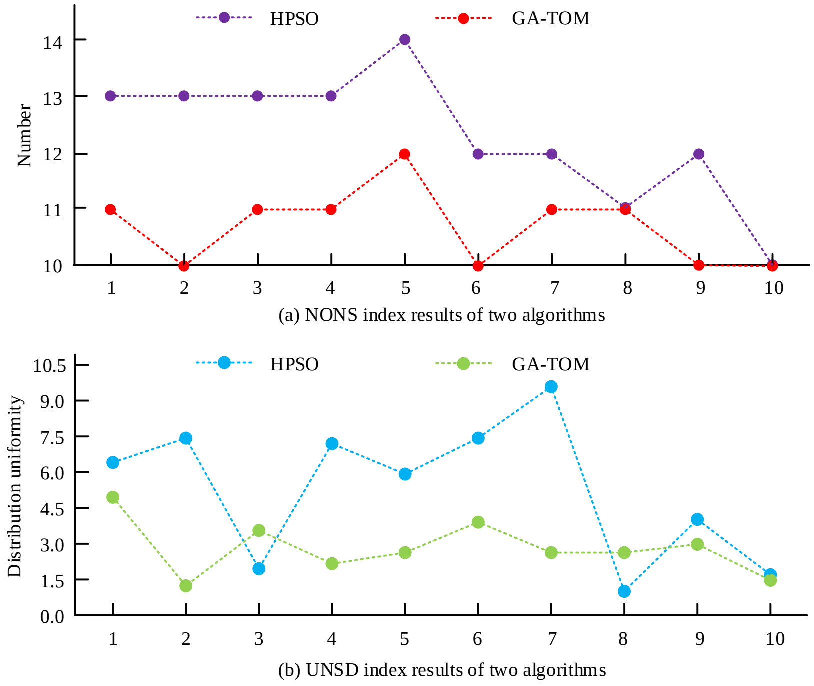
Fig. 7. SCIMC index change curves of four algorithms.

Fig. 7 shows the change curve of SCIMC index of different algorithms. The lower the SCIMC calculation result of the algorithm, the more robust its performance. As can be seen from Fig. 7, the HPSO has a maximum value (MV) of 95.81, a minimum value (IV) of 85.19, and an average value of 89.53. PSO, ACO, and GA-TOM algorithms have MVs of 136.77, 105.45, and 101.37, respectively, the IVs are 86.28, 85.37, and 85.24, respectively, and the average values are 99.09, 90.45, and 89.78, respectively. In summary, the outcomes of the HPSO and the GA-TOM differ less, and the calculation results of HPSO and GA-TOM are lower than the basic PSO and ACO, which proves that HPSO and GA-TOM have better performance in these four algorithms are superior in performance. To compare the performance of HPSO in SCIMC model more scientifically, the study set the machine interference time windows of A $[100,125]$ and B $[125,150]$ respectively, and conducted 10 simulation experiments using GA-TOM algorithm as the control group. Performance indicators include the Number Of Non-inferior Solutions (NONS), uniformity of non-inferior solution distribution, UNSD), dominant proportion of non-inferior solution (DPNS), average distance between non-inferior solution and optimal Pareto (DPNS), ADNSOP), maximum distance between non-inferior solution and optimal Pareto (MDNSOP), Noninferior Solution Dispersion, NSD) and Noninferior solution coverage (NSC), in which the larger the values of NONS, DPNS and NSC, the better the performance, and the smaller the values of UNSD, ADNSOP, MDNSOP and NSD, the better the performance.

Fig. 8 illustrates the experimental results of NONS and UNSD for the two algorithms under a working condition. Fig. 8(a) illustrates the NONS results of the two algorithms, the MV of HPSO algorithm is 14, the IV is 10, and the average is 12.3. The GA-TOM algorithm's MV is 12, the IV is 10, and 10.7 is the average. Fig. 8(b) illustrates the UNSD results of the two algorithms, the MV of HPSO algorithm is 9.700, the IV is 1.454, and the average is 5.283; it is higher than the MV of 4.970, the IV of 1.367, and the average of 2.435 for the GA-TOM algorithm.

The NONS and UNSD variation curves of the two algorithms under B working condition are presented in Fig. 9. The NONS variation curves for the two algorithms are displayed in Fig. 9(a). The maximum, minimum and average values of the HPSO algorithm are 12, 9 and 10.4, respectively, and the values of the GA-TOM algorithm are 11, 8 and 9.3, respectively. Fig. 8(b) shows the UNSD variation curves of the two algorithms, the MV of the HPSO algorithm is 10.500, the IV is 2.526 and the average is 6.132, The MV of GA-TOM algorithm is 12.26, the IV is 1.769, and the average is 5.936. Combining the results of Fig. 8 and Fig. 9, HPSO algorithm is marginally superior to GA-TOM algorithm in terms of the number of non-inferior solutions.

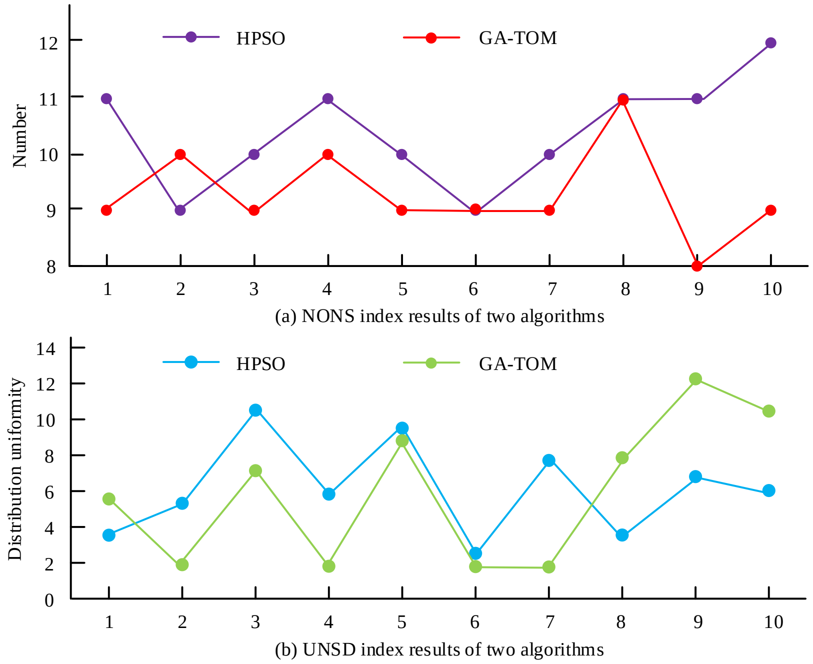
The CM, Dav and Dmax experimental results of the two algorithms under working conditions A and B are shown in Table I. As can be seen from Table I, the experimental results of DPNS, ADNSOP and MDNSOP for the two algorithms under A and B working conditions are presented in Fig. 1. The average of DPNS for the HPSO algorithm is 0.264 and that for the GA-TOM algorithm is 0.069. The average of ADNSOP for the HPSO algorithm is 4.474 and that for the GA-TOM algorithm is 4.485. The average value of MDNSOP for the HPSO algorithm is 4.627 and that for the GA-TOM algorithm is 4.638. The average values of DPNS, ADNSOP, and MDNSOP for the HPSO algorithm are 0.114, 3.104, and 3.189, respectively, for the B condition, and the average values of DPNS, ADNSOP, and MDNSOP for the GA-TOM algorithm are 0.066, 3.110, and 3.193, respectively. In terms of the link between the non-inferior solution set dominance and the separation of non-inferior solutions from the ideal Pareto front, the HPSO method surpasses the GA-TOM.



(a) NONS index results of two algorithms

(b) UNSD index results of two algorithms

Fig. 8. NONS and UNSD results of two algorithms under ‘A’ working condition.



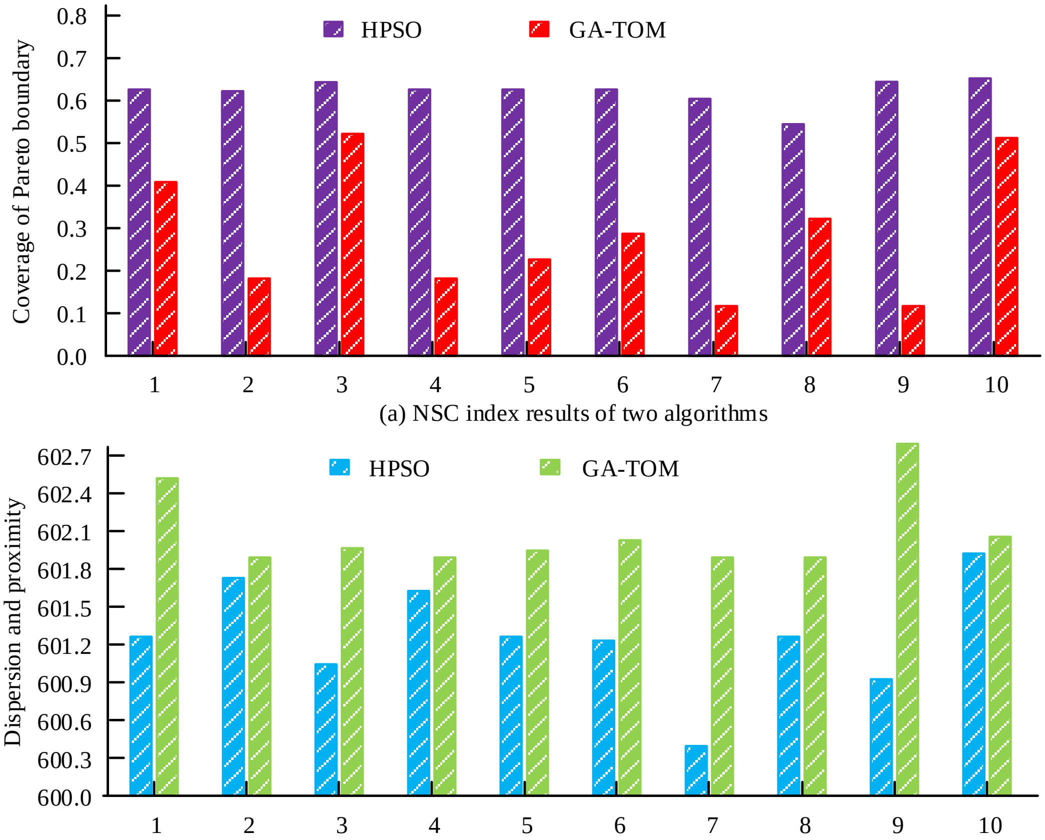
(a) NONS index results of two algorithms

(b) UNSD index results of two algorithms

Fig. 9. NONS and UNSD results of two algorithms under ‘B’ working condition.

TABLE I. DPNS, ADNSOP AND MDNSOP RESULTS OF TWO ALGORITHMS UNDER ‘A’ AND ‘B’ WORKING CONDITIONS

Working condition	Performance index	Algorithm	Maximum	Minimum	Mean
A	DPNS	HPSO	0.814	0.112	0.264
		GA-TOM	0.398	0.000	0.069
	ADNSOP	HPSO	8.028	2.497	4.474
		GA-TOM	8.031	2.512	4.485
	MDNSOP	HPSO	8.264	2.595	4.627
		GA-TOM	8.278	2.611	4.638
B	DPNS	HPSO	0.401	0.000	0.114
		GA-TOM	0.239	0.000	0.066
	ADNSOP	HPSO	3.744	2.601	3.104
		GA-TOM	3.748	2.610	3.110
	MDNSOP	HPSO	3.845	2.681	3.189
		GA-TOM	3.849	2.683	3.193

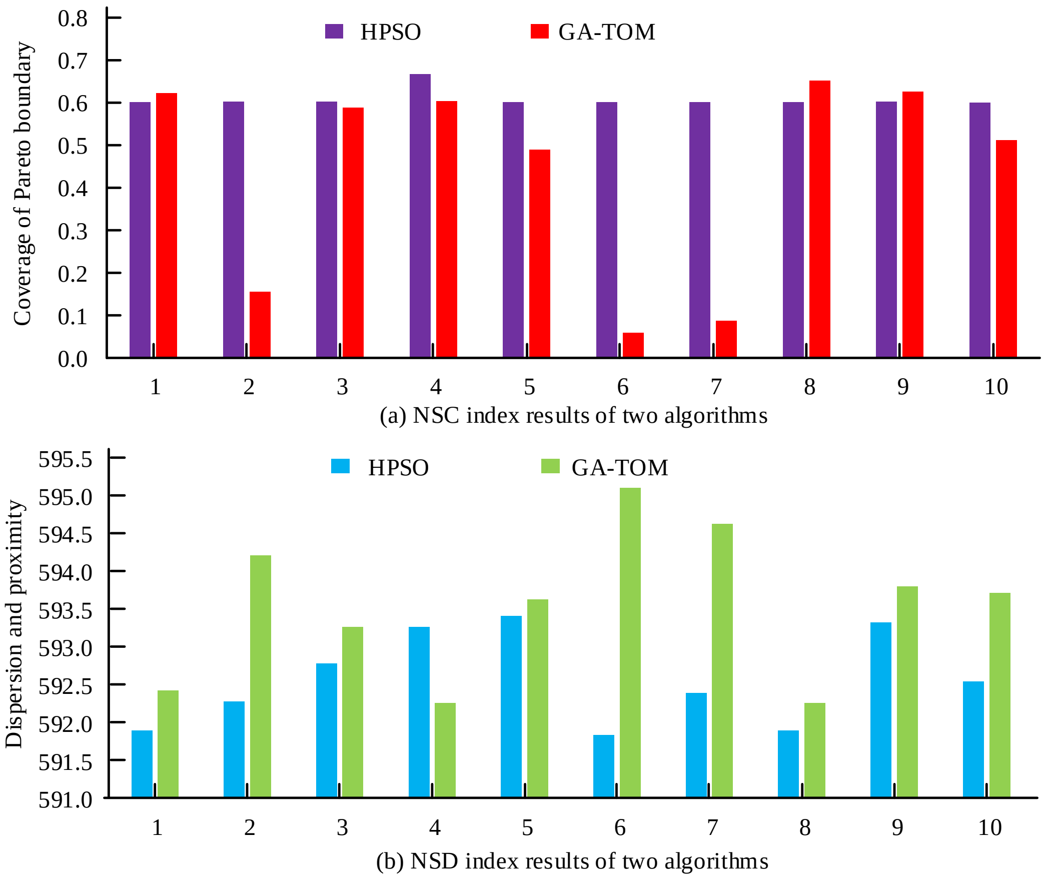


(a) NSC index results of two algorithms

(b) NSD index results of two algorithms

Fig. 10. NSC and NSD results of two algorithms under ‘A’ working condition.

Fig. 10 shows the experimental results of NSC and NSD for both algorithms under ‘A’ working condition. Fig. 10(a) shows the NSC results of both algorithms, the MV of HPSO algorithm is 0.638, the IV is 0.570, and the average is 0.611. The MV of GA-TOM algorithm is 0.511, the IV is 0.124, and the average is 0.298, The IV is 600.581 and the average is 601.300; it is lower than the MV of 602.753, the IV of 600.849 and the average of 601.961 of GA-TOM algorithm.



(a) NSC index results of two algorithms

(b) NSD index results of two algorithms

Fig. 11. NSC and NSD results of two algorithms under 'B' working condition.

TABLE II. COMPARES THE RESULTS OF EACH INDEX OF THE PROPOSED ALGORITHM AND THE TRADITIONAL ALGORITHM IN THE ACTUAL SUPPLY CHAIN INTERFERENCE MANAGEMENT SCHEDULING PROCESS

Evaluation index	Traditional algorithm	Research and propose algorithms
Response time	48 hours	24 hours
Cost saving	5%	12%
Customer satisfaction	75%	88%
Robustness	Intermediate	High
Expandability	Finitude	Good
Innovativeness	There is no	There are

The NSC and NSD variation curves of the two algorithms under B working condition are presented in Fig. 11. The NSC variation curves for both techniques are demonstrated in Fig. 11(a), the maximum, minimum and average values of HPSO algorithm are 0.667, 0.600 and 0.611, respectively. The values of GA-TOM algorithm are 0.652, 0.059 and 0.440, respectively. Fig. 11(b) demonstrates the NSD variation curves of both algorithms, the MV of HPSO algorithm is The MV of the HPSO algorithm is 593.407, the IV is 591.833, and the average is 592.54. The MV of the GA-TOM algorithm is 595.101, the IV is 592.251, and the average is 593.524. Combining the Fig. 10 and Fig. 11, the HPSO algorithm outperforms the GA-TOM algorithm in terms of non-inferior solution coverage, dispersion, and approximation. Finally, in order to carry out the practical application effect of the proposed algorithm, it is applied to the actual supply chain interference management scheduling process. In order to comprehensively evaluate the application effect of the proposed algorithm in the actual supply chain interference management and scheduling process, this study adopts response time, cost saving, customer satisfaction, robustness, scalability and innovation as evaluation indicators. The results of each indicator of the proposed algorithm and the traditional algorithm in the actual supply chain interference management and scheduling process are shown in Table II.

It can be clearly seen from Table II that the proposed algorithm has significant advantages compared with traditional algorithms in the actual supply chain interference management scheduling process. First, in terms of response time, the new algorithm is able to react within 24 hours, while the traditional algorithm takes 48 hours, which indicates that the new algorithm has a faster reaction speed. Secondly, in terms of cost savings, the new algorithm achieved a cost savings of 12%, much higher than the traditional algorithm of 5%, showing higher economic benefits. In addition, the new algorithm also showed a significant improvement in customer satisfaction, reaching 88 percent compared to 75 percent for the traditional algorithm, indicating that the new algorithm was better able to meet customer needs. In addition, the proposed algorithm also shows strong advantages in robustness, scalability and innovation, and has high stability and adaptability. Therefore, it can be concluded that the proposed algorithm has excellent performance and wide application prospects in the actual supply chain interference management scheduling process.

V. Discussion

With the deepening of globalization and networking, supply chain has become the core component of modern enterprise operation. The complexity, dynamics and uncertainty of supply chain bring unprecedented challenges to enterprises. Especially in recent years, due to the global epidemic, trade war, natural disasters and other multiple factors, the stability of the supply chain has been seriously threatened, resulting in a series of problems such as rising enterprise costs, delayed delivery, and decreased customer satisfaction. In order to cope with these challenges, supply chain interference management has gradually attracted the attention of enterprises and academia. SCDM aims to ensure supply chain continuity and stability by identifying, assessing, preventing and responding to disruptive events in the supply chain. However, the traditional supply chain optimization methods are often powerless in the face of complex and dynamic interference events. Therefore, how to effectively manage and dispatch the interference events in the supply chain has become an urgent problem in the field of supply chain management. This study aims to build an efficient and flexible supply chain interference management scheduling model by introducing HPSO algorithm. HPSO algorithm combines the advantages of particle swarm optimization algorithm and other optimization techniques, and can achieve global optimization and fast convergence in complex and dynamic environments. By applying HPSO algorithm, this study is expected to provide a new solution and method for supply chain interference management, help enterprises improve the stability and efficiency of supply chain, reduce operating costs, and enhance customer satisfaction.

The performance of HPSO algorithm in SCIMC model is analyzed and verified by experiments. Firstly, the variation degree of CWB under different influencing factors was analyzed. The results show that job satisfaction is the most influential factor on CWB, while the influence of organizational culture building level is relatively small. This finding is similar to the research results obtained by Bilandi's team in 2021, and this result provides a valuable reference for optimizing SCIMC model, suggesting that more attention should be paid to improving job satisfaction in practical applications [23]. Secondly, through independent numerical experiments on HPSO algorithm, GA-TOM algorithm, basic PSO algorithm and ACO algorithm, it is found that the results of HPSO algorithm and GA-TOM algorithm have little difference, and are better than the basic PSO algorithm and ACO algorithm. This shows that HPSO algorithm and GA-TOM algorithm have higher stability and efficiency in solving SCIMC model. The above results coincide with the research results of XX et al. on HPSO algorithm in 2022 [24]. In order to investigate the performance of HPSO algorithm more scientifically, the research also sets up the machine interference time window in A condition and B condition, and carries out the simulation experiment. The experimental results show that HPSO algorithm is slightly better than GA-TOM algorithm in terms of the number of non-inferior solutions. HPSO algorithm also shows some advantages in the dominant relation of the non-inferior solution set and the distance of the non-inferior solution from the optimal Pareto front. The research results are similar to the performance test results of the improved HPSO algorithm conducted by Zhang's team in 2020 [25].

In summary, the experimental results of this study verify the superior performance of HPSO algorithm in multiple performance indicators, providing strong support for the optimization of non-inferior solution coverage, dispersion and approximation, etc., and the conclusions obtained in this study are also consistent with the conclusions of the latest research. In future research, it is necessary to further explore the application potential of HPSO algorithm in other optimization problems, and constantly improve and improve the algorithm to improve its solving efficiency and stability. At the same time, we will also pay attention to the influence mechanism of key factors such as job satisfaction on the degree of CWB change, with a view to providing more targeted suggestions and guidance for solving practical problems.

VI. Conclusion

With the increasing growth of the economy, the market competition model has undergone a new change, and the traditional competition of enterprise units has been converted into the competition of SC units. Effective SC management can bring more economic benefits to enterprises, but since the SC itself is a dynamic and complex system, its internal is susceptible to disruptive events and the deterioration effect brought by CWB, which leads to hindering the production operation of enterprises. To address this difficulty, the study uses the HPSO algorithm to solve the behavior-based SCIMC model. The experimental results show that the HPSO algorithm for SCIMC objective calculates the MV of 95.81, the IV of 85.19, and the average value of 89.53, which is less difficult than the GA-TOM algorithm with the MV of 101.37, the IV of 85.24, and the average of 89.78. The average of NONS, DPNS, and NSC for the HPSO algorithm under ‘A’ working condition are 12.3, 0.264 The average of ADNSOP, MDNSOP, and NSD for the HPSO algorithm are 4.474, 4.627, and 601.300, which are lower than those of 4.485, 4.638, and 601.961 for the GA-TOM algorithm. The average of ADNSOP, MDNSOP, and NSD of HPSO algorithm are 3.104, 3.189, and 592.54, which are lower than 3.110, 3.193, and 593.524 of GA-TOM algorithm. In summary, the HPSO algorithm proposed in this study has robust performance, can effectively solve SCIMC problems, and promote the development of supply chain scheduling. However, the research still needs to be deepened, especially in considering the multi-benefit objectives of each node enterprise and the complex and changeable negotiation scheduling process. Looking forward to the future, the research in this field can be expanded in the aspects of multi-objective optimization, dynamic scheduling, application of game theory, integration of big data and artificial intelligence, and practical application verification, so as to reveal the internal law of supply chain interference management scheduling more comprehensively, and provide enterprises with more targeted and practical supply chain management strategies. Through these forward-looking discussions and practices, it is expected to promote the research of supply chain interference management scheduling to a new height.

FUNDINGS

The research is supported by: Jiangsu University Philosophy and Social Science Research Project ——Research on the ecological development path of Jiangsu cultural tourism industry from the perspective of "double carbon" (2022SJYB0678).

REFERENCES

[1] Welborn C, Bullington K, Abston K. Recruiting students to an undergraduate supply chain management program. industry & higher education, 2022. Industry & higher education, 2022, 36(2): 190-199.

[2] Zan J. Research on robot path perception and optimization technology based on whale optimization algorithm. Journal of Computational and Cognitive Engineering, 2022, 1(4): 201-208.

[3] Saha S, Chakrabarti T. Cost Minimization Policy for Manufacturer in a Supply Chain Management System with Two Rates of Production under Inflationary Condition. Jurnal Teknik Industri, 2020, 21(2): 200-212.

[4] Wang L L, Liu Z, Zheng Y L, Gu F J. A Two-Sided Matching Method for Green Suppliers and Manufacturers with Intuitionistic Linguistic Preference Information. recent advances in computer science and communications, 2021, 14(8): 2507-2517.

[5] Salehi V, Salehi R, Mirzayi M, Faezeh A . Performance optimization of pharmaceutical supply chain by a unique resilience engineering and fuzzy mathematical framework. human Factors and Ergonomics in Manufacturing, 2020, 30(5): 336-348.

[6] Serenko A. Knowledge sabotage as an extreme form of counterproductive knowledge behavior: the perspective of the target. journal of knowledge management, 2020, 24(4): 737-773.

[7] Choudhury M, De S K, Mahata G C. Inventory decision for products with deterioration and expiration dates for pollution-based supply chain model in RAIRO - Operations Research, 2022, 56(1): 475-500.

[8] Keshavarz T, Noormohamadzade Z, Fakhrzad M B. INTEGRATING PRODUCTION SCHEDULING, DELIVERY, AND 3D LOADING PROBLEMS IN A TWO-STAGE SUPPLY CHAIN. International Journal of Industrial Engineering, 2020, 27(6): 906-932.

[9] Li W, Han D, Gao L, Li X, Li Y. Integrated Production and Transportation Scheduling Method in Hybrid Flow Shop. Chinese Journal of Mechanical Engineering, 2022, 35(1): 112-131.

[10] Goli A, Alinaghian M. A new mathematical model for production and delivery scheduling problem with common cycle in a supply chain with open-shop International Journal of Manufacturing Technology and Management, 2020, 34(2): 174-187.

[11] Solina V, Mirabelli G. Integrated production-distribution scheduling with energy considerations for efficient food supply chains [J]. Procedia Computer Science, 2021, 180(11): 797-806.

[12] Du J, Dong P, Sugumaran V. Dynamic Production Scheduling for Prefabricated Components Considering the Demand Fluctuation. intelligent Automation and Soft Computing, 2020, 26(4): 715-723.

[13] Lee J, Marla L, Jacquillat A. Dynamic Disruption Management in Airline Networks Under Airport Operating Uncertainty. transportation Science. 2020, 54(4): 973-997.

[14] Jiang Y, Ding Q, Wang Y, Junhu R. AN ATTITUDE-BASED MODEL OF DISRUPTION MANAGEMENT TO HANDLE DELIVERY DELAY. journal of nonlinear and convex analysis, 2019, 20(6): 1117-1125.

[15] [Ning T, Duan X, An L, Gou T. Research on disruption management of urgent arrival in job shop with deteriorating effect. journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2021, 41(1): 1247-1259.

[16] Pandi R R, Song G H, Nagavarapu S C. Dauwels J. Disruption Management for Dial-A-Ride Systems. IEEE Intelligent Transportation Systems Magazine, 2020, 12(4): 219-234.

[17] Malik A I, Sarkar B. Paper Disruption management in a constrained multi-product imperfect production system. journal of manufacturing systems. 2020, 56: 227-240.

[18] Abdallah K S, El-Beheiry M M. A SYSTEM DYNAMICS MODEL ASSESSING THE SUSTAINABILITY OF THE PERFORMANCE OF SUPPLY CHAINS WITH REVERSE FLOW. International Journal of Industrial Engineering, 2022,29(4): 562-577.

[19] [Babamiri M, Heydari B, Mortezapour A. Tamadon TM. Investigation of Demand-Control-Support Model and Effort-Reward Imbalance Model as Predictor of Counterproductive Work Behaviors. safety and health at work, 2022, 13(4): 469-474.

[20] Zhang S, Cao L, Lu Z. An EOQ inventory model for deteriorating items with controllable deterioration rate under stock-dependent demand rate and non- Journal of Industrial and Management Optimization, 2022, 18(6): 4231-4263.

[21] Jain K, Saxena A. Simulation on supplier side bidding strategy at day-ahead electricity market using ant lion optimizer [J]. Journal of Computational and Cognitive Engineering, 2023, 2(1): 17-27.

[22] Abdelghany M, Yahia Z, Eltawil A B. A new two-stage variable neighborhood search algorithm for the nurse rostering problem. RAIRO - Operations Research, 2021, 55(2): 673-687.

[23] Bilandi N, Verma H K, Dhir R. hPSO-SA: hybrid particle swarm optimization-simulated annealing algorithm for relay node selection in wireless body area networks. Applied Intelligence, 2021, 51(5): 1410-1438.

[24] Ramasamy V, Thalavai Pillai S M. An effective HPSO-MGA optimization algorithm for dynamic resource allocation in cloud environment. Cluster Computing, 2020, 23(6): 1711-1724.

[25] Zhang P, Hong Y, Pang X, Jiang C. VNE-HPSO: Virtual network embedding algorithm based on hybrid particle swarm optimization. IEEE Access, 2020, 8(3): 213389-213400.

Presentation of a New Method for Intrusion Detection by using Deep Learning in Network

Hui MA*
Modern Educational and Technological Center, Henan Quality Institute Pingdingshan, Henan, 467000, China

*Abstract***—Intrusion detection in cyberspace is an important field for today's research on the scope of the security of computer networks. The purpose of designing and implementing the systems of intrusion detection is to accurately categorize the virtual users, the hackers and the network intruders based on their normal or abnormal behavior. Due to the significant increase in the volume of the exchanged data in cyberspace, the identification and the reduction of inappropriate data characteristics will play a significant role in the increment of accuracy and speed of intrusion detection systems. The most advanced systems for intrusion detection are designed for the detection of an attack with the inspection of the full data of an attack. It means that a system of detection will be able to recognize the attack only after the execution of the attack on the attacked computer. In this paper, a system for end-to-end early intrusion detection is presented for the prevention of attacks on the network before these attacks cause further detriment to the system. The proposed method uses a classifier based on the network of the deep neural for the detection of an attack. The proposed network on a supervised method is trained for the exploitation of the related features by the raw data of the traffic of the network. Experimentally, the proposed approach has been evaluated on the dataset of NSL-KDD. The extensive experiments show that the presented approach performs better than the advanced approaches based on the accuracy, the rate of detection and the rate of the false positive, and also, the proposed system betters the rate of detection for the classes of the minority.**

Keywords—Attack; security on cyberspace; classification; intrusion detection; deep learning

I. Introduction

The ever-increasing expansion of the Internet, in terms of infrastructure and in terms of software, has caused an increment in the number of network users' number and their applications [1]. Today, many public sector services and many private sector services are virtually done on the Internet. The development of this virtual space has caused the detection of intrusion to become the most important subject in the scope of the security of computer networks. The systems for intrusion detection try to classify the activity of the made connections by the users into two categories: the normal and the abnormal [2]. In more advanced systems, sometimes, the type of abnormal behavior, which is also called an attack, is specified. Each connection in the network is described based on the collection of features, which these features can be used for the determine of the normal connection or the abnormal connection [3], [4].

The accuracy of the detection for the systems of intrusion detection is the most important indicator of the efficiency of these systems. The increment of the accuracy in the systems of the detection of the intrusion prevents the result of more attacks in the system. The attack neutralization will play a decisive role in the reduction of the costs caused by the attacks on valuable network resources [5]. The attacks of the Denial of Service (DoS) are the most popular kind of attacks that are constantly sent to servers by different IP addresses and prevent the serving of the servers or shut down of the servers [6]. Usually, the attacker varies the address of the IP to carry out an attack on the victim's computer. Although the network firewall that first responds to the attack performs the process of filtering the attack, unfortunately, today, this amount of security is not enough. Then, the attacks that can elapse the firewall are sent through an intermediate router into the deep learning-based approach to the detection of the intrusion. Here, the multitude of variant classes of new attacks can be identified. If an unusual position is observed, it is immediately related to the center of the management of the system. Usually, this center informs the system officers about this situation via SMS or email, and then it automatically blocks this attack [7], [8].

It is essential to provide a more precise method of the detection of attack for the detection of novel attacks in the servers and the networks. The detection methods do not use the programming of the unequivocal according to problem complexity [9]. Usually, they use the methods of machine learning, which can be remarkably prosperous on the problems of decision-making as long as the features of the sufficient are presented on problem [9], [10]. In deep learning, incumbent features are also exploited by training certain layers like Recurrent Neural Networks (RNNs) or Convolutional neural networks (CNNs). The achievement of an extended network of deep learning strongly appertains the kind of layers of feature extraction. Also, it depends on sufficiently large training dataset. In addition, the preprocessing and organization of input data may have an important result in achievement [11]. Thus, hunt for superior methods in different fields is a topic that is studied by researchers. In current paper, a system for end-to-end early intrusion detection is presented for prevention of attacks on network before these attacks cause further detriment to the system. The proposed method uses a classifier based on the network of the deep neural for the detection of an attack. The proposed network on a supervised method is trained for the exploitation of the related features by the raw data of the traffic of the network.

In summary, the contributions and the reasons for choosing the methods of this paper are as follows: 1) A basic network intrusion detection method is presented. If more data is available, the proposed approach can make a more informed decision about the class label of a given network traffic data, but it delays the decision. Therefore, in this work, the focus is on optimizing the attack detection accuracy with minimum delay. 2) The proposed approach extracts relevant features from raw network traffic data end-to-end instead of relying on manual feature engineering process. Therefore, the proposed approach is domain-independent and does not require domain-specific data preprocessing steps. 3) A new metric is introduced to evaluate how early the proposed approach can detect attacks.

The continuation of this article is as follows: Section II provides a review of the research literature. Section III the proposed method, which uses the model of deep learning, is presented. In Section IV, the dataset used in the designed experiments details and the outcomes of these experiments are presented. In Section V, the general conclusions and the prospects for future research are presented.

II. Review of Literature

The system of intrusion detection in the network is applied for the monitoring of the network traffic to protect the system from network threats. This system tries to find the destructive acting like the attacks of DoS, the attacks that monitor the network traffic, and the port scanning attacks [12]. In general, two methods are used to detect the intrusion on the network:

A. Detection of Anomaly

It is when the observed behavior of the user does not follow an expected behavior [12]–[14]. In network anomaly detection, the normal system activities such as the network bandwidth, the ports, the rules and the device communication are examined. The detection of the anomaly of the intrusion is a hard problem with several proofs. First, the use of the system and user behavior is constantly evolving. Therefore, the intrusion detector must also evolve. Without the permission for these changes in the behavior, the network administrator will soon be inundated with false alarms, and it will rapidly affect the trust in the system. A second important factor in the detection of the anomaly is that an alert of the behavior of the abnormal may not furnish any specific beneficial data for the administrator of the network. The ambiguous alert about which the system may be beneath the attack makes it hard to catch the firm's measured action. Also, sometimes, the known attack may not be recognized by the system of the detection of the anomaly. This is the fundamental problem [12]– [14].

B. Signature-based Detection

It is when the observed behavior indicates an intention for the misuse of the network computing resources. The benefits of the detection based on the signature consist of simpleness and effectiveness, and it has a great ability for detecting the attacks of the known. The other important profit is that the alert that is published is a certain alert because it reconciles the signature of the pattern of a certain attack. By a special alert, the administrator of the network can rapidly assess whether the attack of the doubtful is a wrong alert or real. I.e., if it is detected as real, the network administrator can adopt an appropriate response. Another advantage of this approach is the ability to produce accurate results and reduce false alarms. However, a disadvantage of the system of detection based on signature is that the signature file must be updated. Also, with the increment of the number of signatures, the efficiency of the system is decreased. On the other hand, the system will be able to recognize the attacks of the known. The smallest change in the known attacks causes the possible loss of the attack detection by the system [15]–[17].

In the literature, there are different researches for IDS with the use of the methods of machine learning. Here, the latest articles based on the deep learning are presented. In [18], the authors have presented a hybrid method based on the networks of the deep belief and the networks of the probabilistic neural. The authors deal with the imbalance of the class in the NSL-KDD dataset with the use of SMOTE about the increment of the classes of the minority and with the use of NCL about the classes of the majority of the under-sampling. In [19], the authors have presented a technique for the extraction of the feature with the use of the sparse auto-encoder. Their presented system is contrasted to the prior methods of the extraction of the feature. The process of the classification is speeded up, and a superior process for the learning is acceded. An impressive practical model has been developed for use in the systems of the detection of intrusion.

In study [20], the authors have proposed a revision of the literature on machine learning and the applications of deep learning in the detection of intrusion on the network. Also, the authors have appraised the different databases, the different approaches and the different problems with the detection of the intrusion. In research [21], the authors have provided the models based on the optimization of the particle swarm for the selection of the feature and for the meta-parameter selection. They have evaluated the different models of deep learning, like the Long Short-Term Memory (LSTM) and the Deep belief network (DBN). In study [22], the researchers have provided a review of research about deep learning for the IDS. They have also reviewed the light research and the directions for the future. The authors provide an approach for the architectures of deep learning for the kinds of IDS and for the databases. In research [23], the authors have proposed the categorization of the models of deep learning for the systems of the detection of intrusion. Also, they have presented the literature review in their research. The training and the test have been performed on the various databases with the use of $four$ methods of deep learning. Their test outcomes are compared by articles in the literature. Also, the assessments for the latter research about the systems of the detection of the intrusion, which are based on deep learning, are provided.

The authors have provided an approach to the detection of the attack based on deep learning about the attacks of Distrubed Denial of Service (DDOS). Their presented approach has been evaluated in the dataset of CICIDS. Also, it has been simulated the traffic of the DDOS. It is expressed that the presented approach outperforms the several prior approaches for the detection of the attacks of the DDOS [24]. In [25], the authors have provided an asymmetric deep auto-encoder classification method for the unsupervised learning of the features. Their proposed method is evaluated on GPU with the use of the dataset of KDD-CUP99 and the dataset of NSL-KDD. The obtained outcomes have been contrasted by the research on literature. In study [26], the researchers have extended a model of the deep network, which consists of the

RNNs with the units of the recurrent of the gated, the Multi Layer Perceprton (MLP) and the modules of the softmax for an increment of the efficiency of the systems of the detection of the intrusion. Their experiments on the presented model have been performed in the dataset of KDD-99 and the dataset of NSL-KDD. The outcomes of the experiments have displayed that the GRU has superior outcomes over LSTM in the systems of the detection of intrusion. In study [27], the authors have extended an IDS basis on self-learning according to the framework of STL to propose the superior security of the network over the standard technologies for the defense of the network. The experiments in the dataset of NSL-KDD incur accuracy of the classification of binary and the accuracy of the five-class classification. Also, their experiments reduce the time of the training and the time of the test.

In research [28], the authors have provided an approach based on the cloud for the detection of the intrusion in the network in real-time with the use of the binomial deep learning models and the models of the 5-class deep learning along with the framework of H2O. In case of an attack, the models can dispatch a notification to the mobile to the model authorities with the help of a web page on the architecture basis on the cloud, which the authors are planning. In study [29], the authors have proposed a model of IDS based on the improved DBN by the algorithms of the genetic for IoTs. The structure of the network of the optimal with GA is characterized by the use of numerous iterations. In study [30], the researchers have investigated a model of DNN by the different layers. Also, authors have used the NSL-KDD dataset, the UNSW-NB15 dataset, the Kyoto dataset, the WSN-DS dataset and the CICIDS 2017 dataset for NIDS and HIDS. Due to the done tests, this approach can be scouted on the time of the real, and it has superior outcomes over the traditional algorithms of machine learning. In research [31], the researchers have extended an approach based on DBN by a layer of the classification with four-layer and Support Vector Machine (SVM). These authors have done the experiments using different kernels like Radial Basis Function (RBF), the linear, the sigmoid and the polynomial. Due to the empirical outcomes in the dataset of NSL-KDD, authors have obtained the foremost outcomes with the use of the kernel of RBF.

III. The Presented Method

In the current section, the proposed system for the detection of intrusion and the detection of network attacks is presented. The primary purpose of the presented method is the monitoring of the traffic of the network on the time of the real, the extraction of the automatic features from the raw data of the traffic of the network, the prevention of the time-consuming task of the feature extraction using the traditional methods and the accurate detection of the attacks in the network. The presented method can be intersected into two general stages. The overview of the presented method is displayed in Fig. 1. In the first step, the proposed flow classifier is trained and then evaluated using a dataset that has the labeled flows of the network and the related packets of the network into these flows. In the second step, the trained classifier of the flow is used for the prediction of whether a given flow from the network is destructive or the usual. The flow of the network is the two-way trail of the packets, which is swapped among two endpoints in a specified interval of time by several joint features of the flow [9], such as the addresses of the IP of the source, the addresses of the IP of the destination, the numbers of the port of the source, the numbers of the port of the destination and the protocol kind. In the proposed method, a flow of the network is defined as a trail from the regular packets T. In it, T displays the longitude of a full flow. The given flow is shown as follows:

$$F_T = \{P_1.P_2.\cdots.P_T\}\ \forall P_i \in \mathbb{R}^d \wedge 1 \leq i \leq T \tag{1}$$

d is the packet length. Two main steps of the proposed method use a flow processing approach, which this approach includes three modules: the filtering of the packet, the identification of the flow and the preprocessing of the packet. The module of the filtering of the packet takes the packets of the network and then sends them into the latter modules if these packets meet certain criteria. The next modules convert the packets, and then these modules categorize them in the flows of the network. When a flow of the network is updated by a novel packet, then the classifier of the flow is used for the updation of the related prediction to the flow.

A. *Approach of the Processing of the Flow*

In the current section, the modules of the approach of the designed flow processing are introduced. As mentioned in the previous section, in this approach, three modules are used, which are as follows:

1) *Filtering of packet:* The traffic of the raw of the network among the unreliable network and the attacked system is monitored. Just the packets from the network are selected that are relevant to the kind of attacks that have to recognize them. For instance, if the goal is the detection of web attacks [10], then we only capture the packets of HTTP.

2) *Identification of flow:* With the reception of a novel packet, the model checks the packet features, like the addresses of the IP of the source and the addresses of the IP of the destination, for the identification of a flow of the appropriate activity for it. The flow of the active displays a session of communication between two endpoints of the network. Also, if no active flow is found that matches the features of the packet, then a novel flow is created. A flow of the network is presumed to be inactive or terminated. After that, the connection is lost or the time at which the flow has not embraced a novel packet on a specified time period. The value of the timeout of the flow can be set due to the kind of protocol of the traffic of the network, which is adapted for the detection of attacks.

3) *Preprocessing of packet:* When the suitable flow of the novel packets is identified, then every packet is sent via the below preprocessing stages to truncate the unfavorable data and to convert it to the byte vector with a uniform size: the truncation step and the transformation step. The primary goal of the mentioned stages is the confidence that the classifier must rely on the related features to classify the flow. The details of these steps are as follows:

a) *Truncation step:* The packets of the raw received consist of the header of Ethernet. This header contains the data about the link of the physical. However, this information is worthless for the detection of an attack insomuch it can be faked as easily. Therefore, the header of Ethernet is deleted from the packet. Also, the header of the IP on the packets contains data like the perfect longitude of the packet, the version of the protocol, the addresses of the IP of the source and the addresses of the IP of the destination. The mentioned data is essential for the routing of the packets on the network. This data is considered disjointed for the classifier because it is possible that the classifier starts to emphasize the IP information for the detection of the flows of the attack. Thus, it is deleted from the packets. It permits the classifier to operate consistently if the node's address on the network has varied. Also, it allows the classifier to popularize the learned information from an environment of the network to another environment.

b) Transformation step: At the time of the usage of the neural network for the classification operation, an input with a fixed size is required. In order to uniform the header longitude of the layer of the transport and in order to uniform the payload of the packets, these packets with the zeros to a fixed length are cutted. It should be noted that if the packet's longitude on a flow is limited, then, the flow longitude is not limited, unlike the other proposed methods like the presented approaches [11].

B. *Flow Classification*

An essential aspect of proper forecasting is the prediction of the attack with enough time to implement the true reaction to the attack as rapidly as possible before it causes further damage to the attacked system. Therefore, to minimize the time of the prediction, a DNN is used as a classifier of the flow insomuch with an increment of the model size, the prediction time is increased because more computations are required. The key problem is that the decrement of the size of the model usually makes the finite power and the little accuracy [32], [33]. To alleviate the mentioned problem, instead of the training of a model of the large complex for the classification of all kinds of attacks, a set from the naive models is trained which in it, every model is trained for the classification with just a subset of the classes of the attack. Several ensemble strategies, like the voting of the majority and the ranking, have been proposed in the literature for the employment of multiple models and the combination of their predictions [34]. However, the discussion of the ensemble strategies is beyond the domain of the current article.

In the current article, the networks of the neural convolutional as one-dimensional [35] are used for the extraction of a good representation of the internal flows of the network and the presentation of it as the input of a network of the fully-connected or the network of the dense. A layer of the softmax [36] is used as the layer of the final of the network for the calculation of the distribution of the probability for the classes of the target. CNNs are used for the extraction of the related features by the data of the input of the network, like the images and the trails. These networks are able to model the dependencies of the spatial and the dependencies of the temporal on the data with the learning of the corresponding filters of the convolution. A convolution layer consists of several convolution filters, and each filter is used for the extraction of a feature of the specific from the data of the input. Therefore, the output of the layer of the convolution is named the map of the feature.

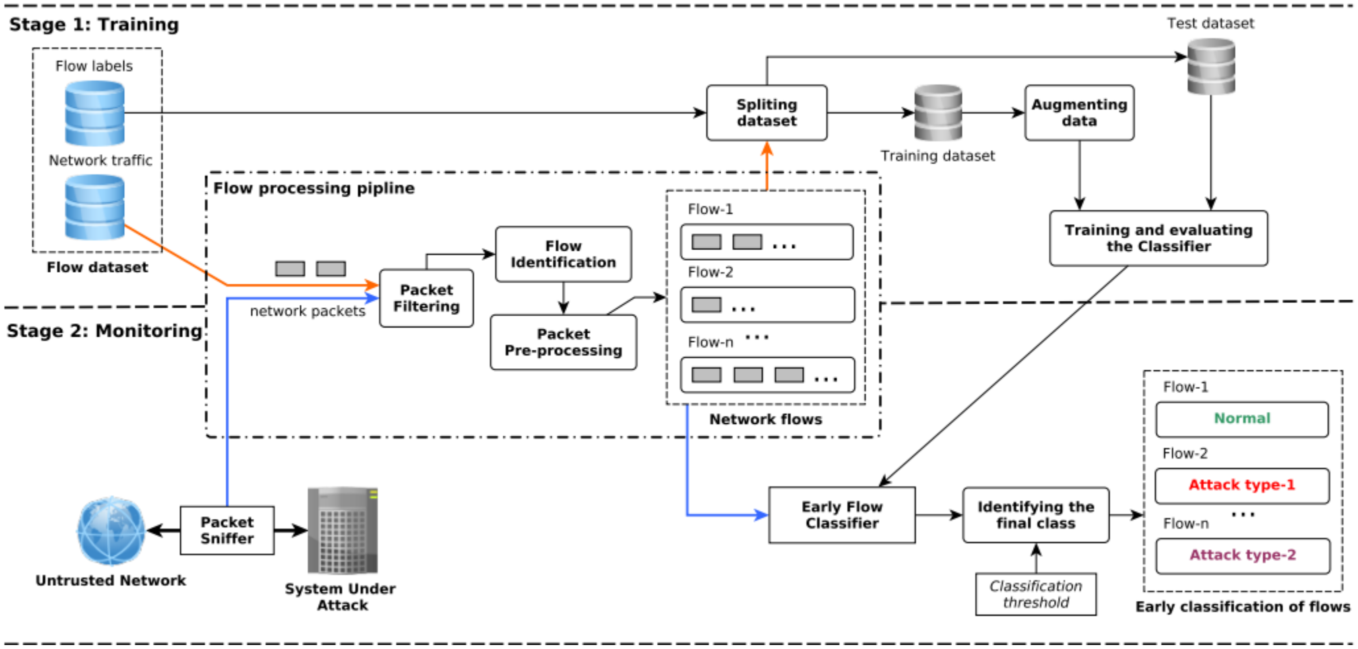


Fig. 1. An overview of the steps of the presented model for network intrusion detection.

The data of the input of ID-CNN has two dimensions. 1-th dimension determines the trail of the circumstances. In contrast, 2 -th dimension is related to the features of the individual from a circumstance. ReLU [37] is used as a function of the activation of the nonlinear in each neuron on the layer of the convolution. Usually, every layer of the convolution is coupled with a layer of the pooling [36] to get the translation-invariance from the returned output with the layer of the convolution. The mentioned layer decreases the size of the time of the output by replacing every partition with the static size by a statistic of the summary from the adjacent elements. CNNs have fewer trainable parameters than the other kinds of ANNs [35]. Therefore, they have fewer possibilities for the overfitting of the data of the training versus the networks of the fully connected. After the convolution operations and the operations of the pooling, a flow of the network with the changeable longitude is displayed with a series of maps of the feature with the changeable longitude. A global layer of the pooling [38] is used for the transformation of the series to a vector with the changeable longitude, and this vector is presented as the input of the layers of the fully connected to obtain the vector of the feature. Finally, a layer of the softmax is applied to the vector of the feature to achieve a distribution of the probability for every class. Based on the distribution of the probability and the basis of the threshold of the classification, the predictions of the final are made.

C. Training

The classifier is trained before its use for the detection of early intrusion online. The system's purpose is to learn automatically the features of the spatial-temporal from the flows of the raw of the network, and then, this system uses these learned features for the reliable identification of the attack flows as quickly as possible. a dataset of the flow labeled is needed for the training of the supervised, which dataset includes the normal flows and the flows of the attack. Moreover, the flows of the labeled and the dataset must contain the corresponding packets of the network with flows. Most of the datasets that are used to train IDSs suffer from the imbalance of the class [2]. That is, the sample number between the various classes in the dataset is not the same. The trained classifier on an unbalanced dataset usually shows the bad efficiency basis the accuracy of the prediction. Thus, in the current article, to correct the effect of the imbalance of the class, the used classifier is trained by the weighting of the sample, which this weighting performs as a factor for the value of the computed loss in every sample among the process of the training. The weight of every sample is based on the related class. Inversely, it is computed with the frequencies of the class in the data of the training of the proportional. The goal is that the classifier must pay further attention to the examples which belong to the classes of the down-represented.

The dataset of the training is prepared with the processing of each packet into flows with the use of the mentioned method above. A dataset of the flow is displayed as $= \{(F_T^{(i)}.y_j)\}$ $1 \leq j \leq N$. In it, N is the flow number F_T and y shows their labels. As regards the goal, it is the training of a classifier that can reliably detect an attack flow after the view of the first packets from a certain flow. Thus, the dataset is extended with the creation of a stack of the short fragments of a flow with the variable longitudes. The expanding process of the dataset of the training with the generation of more data than the existing data is named the augmentation of the data [39]. The process start with the creation of the smallest fragment from a certain flow that contains just the first packet from the flow. Subsequently, by cumulatively adding more packets, more fragments per flow are created, according to the predefined fragmentation rate s_r such that $0 < s_r < 1$. The fragmentation rate s_r is a meta-parameter which is used for the calculation of the fragment size $s_z = [s_r * T]$ in a certain flow, which in it, T is the flow longitude. The value of this parameter reins the number of the produced fragments in every flow. For example, with the decrement of the value of s_r, the more fragments are generated in each flow.

Presume that a flow is given as $F_T^{(j)} = \{P_1.P_2.\cdots.P_T\}$, the fragments set of this flow is as below:

$$\left\{F^{(j)}_{t=k*s_z}\right| k = 1.2.\cdots.\left\lfloor\frac{T-1}{s_z}\right\rfloor \qquad (2)$$

All fragments have the label of a similar y_j which the main flow is needed. For instance, suppose 3 flows by the various longitudes: $F_6^{(1)}$, $F_{15}^{(2)}$, and $F_{70}^{(3)}$. The fragmentation rate s_r is set with 0.25. The fragment sizes s_z for $F_6^{(1)}$, $F_{15}^{(2)}$ and $F_{70}^{(3)}$ are determined $2, 4$ and 18. Table 1 catalogs the fragments from the generated flows with the process of the augmentation of the data.

TABLE I. THE FRAGMENTS FROM THE GENERATED FLOWS BY THE DATA AUGMENTATION PROCESS FOR THE MENTIONED EXAMPLE

No.	Flows	Flow Segments
1	$F_6^{(1)} = \{P_1.P_2.\cdots.P_6\}$	$\{P_1.P_2\}$
2		$\{P_1.P_2.\cdots.P_4\}$
3	$F_{15}^{(2)} = \{P_1.P_2.\cdots.P_{15}\}$	$\{P_1.P_2.\cdots.P_4\}$
4		$\{P_1.P_2.\cdots.P_8\}$
5		$\{P_1.P_2.\cdots.P_{12}\}$
6	$F_{70}^{(3)} = \{P_1.P_2.\cdots.P_{70}\}$	$\{P_1.P_2.\cdots.P_{18}\}$
7		$\{P_1.P_2.\cdots.P_{36}\}$
8		$\{P_1.P_2.\cdots.P_{54}\}$

The data augmentation is applied only to the flows of the dataset of the training. This dataset is expanded with the inclusion of the created flow fragments. Table I shows the fragments from the generated flows by the data augmentation process. The proposed classifier is trained for the learning of the map function $H: F_t^{(j)} \rightarrow y_j$ (where $t \leq T$). Namely, the classifier must be able to predict the label of the class y_j from a certain flow $F_t^{(j)}$ by just the first packets t. In the proposed method, the function of the loss of the cross-entropy and the optimizer of Adam [40] have been used for the training of the proposed classifier.

D. Monitoring

Real-time monitoring on high-speed networks is challenging work because of the great rates of the packet. In the proposed method, the mentioned point is the main reason that only process the packets of the network that are relevant to the attack's kind, which want to recognize. The module of the sniffer of the packet is responsible for the monitoring of the traffic of the network on the time of the real. As shown in Fig. 1, it captures and then forwards the incoming network packets and the outgoing packets of the network to the approach of the processing of the flow. The mentioned module, with the use of the library of libpcap is implemented, which furnishes an interface of the programming for the capturing of the packets which are passed via the interfaces of the network. Also, this library asserts the filters that can be configured to take just the specific packets. These filters, which usually are supported by the kernel of the system of the operating, cure the efficiency with the reduction of the overhead of the process of the filtering of the packet.

Also, a roster of the flows of the active and the prognostications about these flows are holded, which are done by the classifier of the flow. As shown in Fig. 2, while the flow of the network is updated by a novel packet, the classifier of the flow is used for the obtention of a prediction. The class of the final of a flow is the class that has the greater probability over the other classes, and classification threshold $\in [0.1)$. Whenever none of the probabilities of the class is greater than the certain threshold, then the proposed method returns Unknown as the class of the final. If the threshold of the classification is increased, then the rate of false positives is reduced, which improves the classification accuracy. The thresholds are presented by the administrator of the network, who sees the traffic of the network, and this person is responsible for the reaction to the attacks.

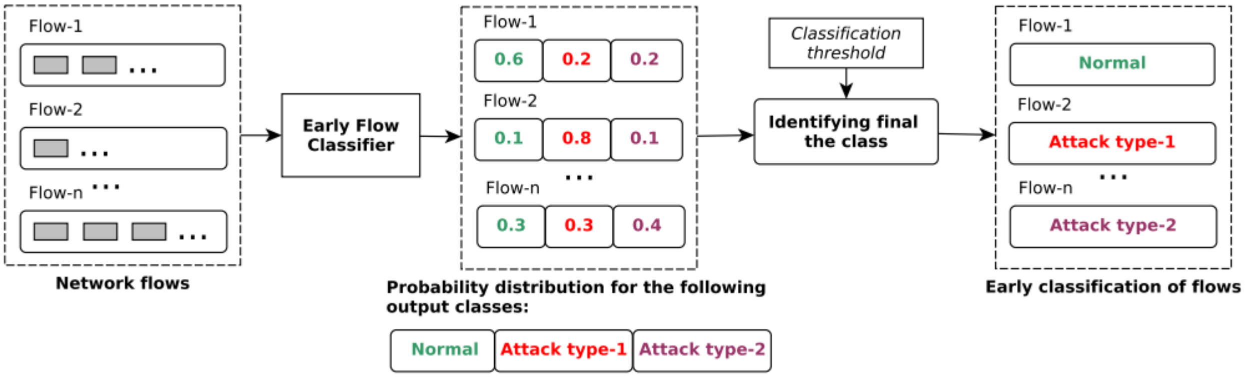


Fig. 2. The flow classification process.

IV. Experiments and Evaluation of Results

In the current part, the used dataset, the performed experiments and the obtained results are presented. The presented method is implemented on the computer by Core (TM) i7 CPU 3.0 GHz Intel(R) and 8G RAM. The efficiency of the presented approach is tested in the dataset of NSL-KDD. The data distribution of this dataset is shown in Fig. 3. It should be noted that the distribution of the data is unbalanced. The efficiency is displayed with the performing of the comparisons by the other deep learning models for the verification of the performance of the various components.

A. The Evaluation Criteria and the Parameter Setting

In the designed experiments, the strategy of K -Fold Stratified cross-validation is used for the test and for the training to ensure a good ratio of the training test. The optimizer of Adam [40] is applied as an optimizer for the optimization of weights for the training. The proposed model is trained in 150 iterations on the architecture of the basic. The rate of learning is adjusted to 0.001. Also, the exponential rate of the decay for the estimation of the first moment is adjusted to 0.9, and the estimation of the second moment is adjusted to 0.999. The tests are done with the use of Keras. Then, the proposed approach is appraised, due to three criteria: FPR, ACC and DR. ACC evaluates the ability of the model to predict the traffic of the normal and the traffic of the attack, while DR measures its ability for the detect of the attack traffic. A higher DR indicates that this approach is susceptible to the attacks of the network and that it aids in the persons to act the precautions on time. FPR is applied for evaluation of the misclassification of the traffic of the normal. DR and FPR should be presumed as common since a great DR can be overshadowed by a great FPR. The above criteria definitions are as follows:

$$ACC = \frac{TN + TP}{FP + FN + TP + TN} \tag{3}$$

$$DR = \frac{TP}{FN + TP} \tag{4}$$

$$FPR = \frac{FP}{TN + FP} \tag{5}$$

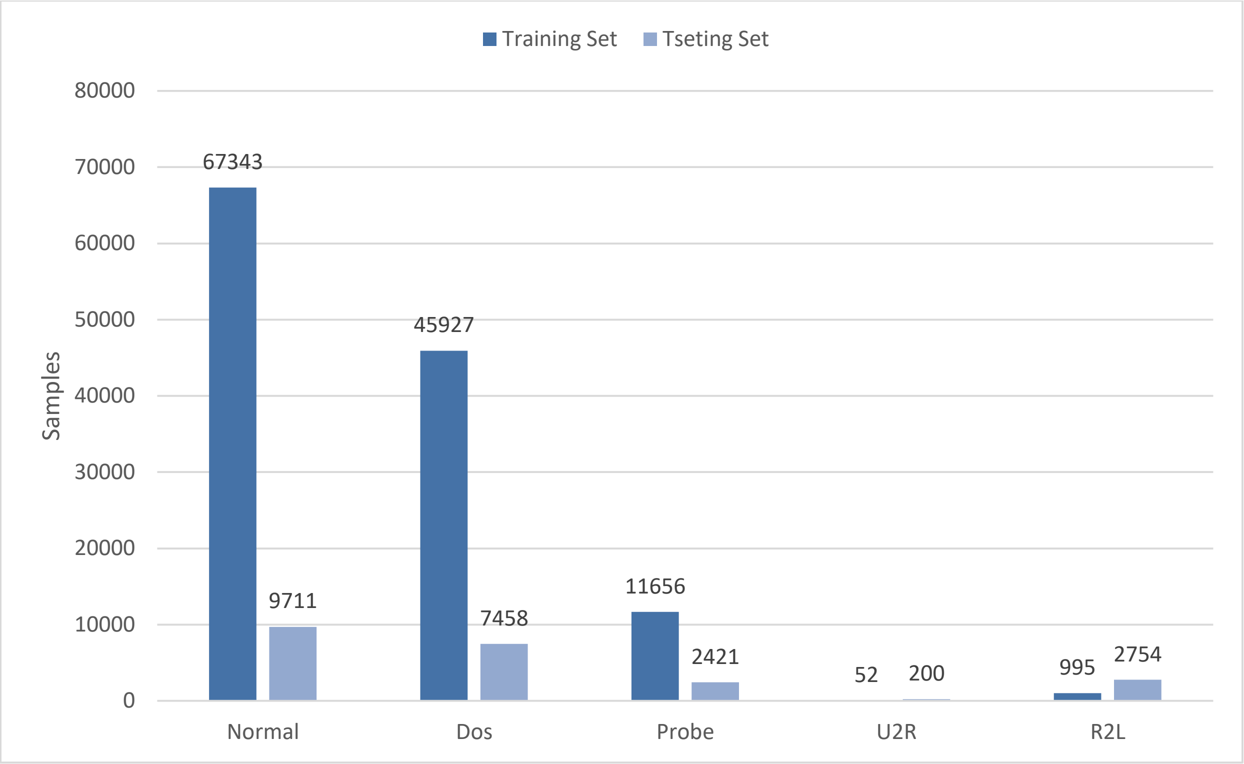


Fig. 3. The data distribution in the dataset of NSL-KDD.

TN and TP display the number of the traffic of the normal and the number of attacks that are properly classified, respectively. FP display the number of the records of the real normal that are misclassified as attacks. Also, FN display the number of attacks that are misclassified as the traffic of the normal.

B. The Acquired Results

To display the performance of the presented approach, two scenarios are considered: 1) the binary classification where the proposed method judges whether an instance is an attack or the traffic of the normal. 2) the classification of the multi-class where the proposed approach forecasts whether an instance is a class from the given attacks on the dataset or the normal.

1) The classification of the Binary: Table II displays the outcomes of the classification of the binary with the K-Fold Stratified cross-validation, where k is in the range of 2-10. For the dataset of NSL-KDD, the mean ACC is equal to 99.71%, the mean DR is equal to 99.75%, and the mean FPR is equal to 0.28%. In Fig. 4, it is clear that the presented method has the great DR and the great ACC. Also, it has a little FPR. With analysis of Fig. 4, it can be seen that the foremost outcomes become visible at K equal to 10. Because with an increment of the folds number, there will be more examples from every class of the attack/normal that the approach can be trained with them. Thus, the approach can classify them as superior. The outcomes show that the proposed method has the powerful ability to discern between the traffic of the normal and the traffic of the attack.

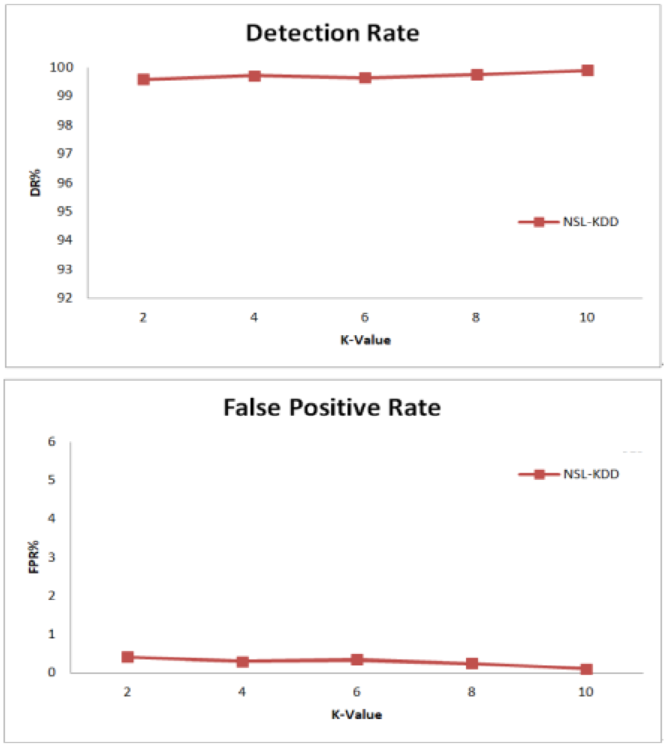


Fig. 4. The comparison of DR and FPR in the binary classification mode

TABLE II. THE OUTCOMES OF THE PRESENTED APPROACH ON THE BINARY CLASSIFICATION MODE

K	ACC	DR	FPR
2	99.62	99.63	0.40
4	99.66	99.73	0.27
6	99.62	99.67	0.36
8	99.81	99.79	0.25
10	99.83	99.93	0.11
Average	99.71	99.75	0.28

2) The classification of the multi-class: The outcomes of the classification of the multi-class are presented in Table III. For the dataset of NSL-KDD, the mean ACC is equal to 99.71%, the mean DR is equal to 99.73%, and the mean FPR is equal to 0.32%. Fig. 5 displays DR of the classes for the five-classes classification in the dataset of NSL-KDD. As it is clear from Fig. 5, the proposed method shows excellent performance in the detection of malicious attacks. To display the outcomes of the detection of the approach directly, a matrix of the confusion is used for the representation of the outcomes of the test. Fig. 6 displays the matrix of the confusion in the dataset of NSL-KDD for the classification of the multi-class. According to Fig. 6, it is shown which most instances are focused on the matrix diameter, and this point shows that the total efficiency of the detection is great.

TABLE III. THE OUTCOMES OF THE PRESENTED APPROACH ON THE MULTI-CLASS CLASSIFICATION MODE

K	ACC	DR	FPR
2	99.58	99.46	0.33
4	99.72	99.85	0.40
6	99.69	99.77	0.37
8	99.81	99.78	0.20
10	99.74	99.80	0.31
Average	99.71	99.73	0.32

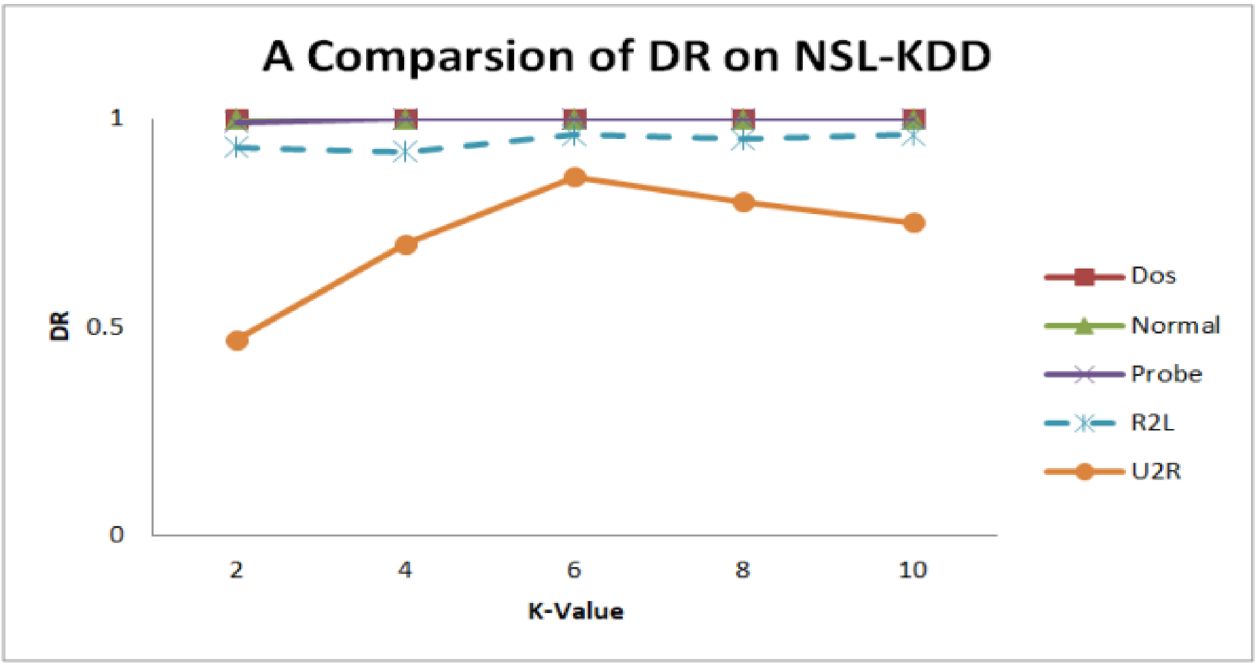


Fig. 5. The DR results of all classes in the dataset of NSL-KDD.

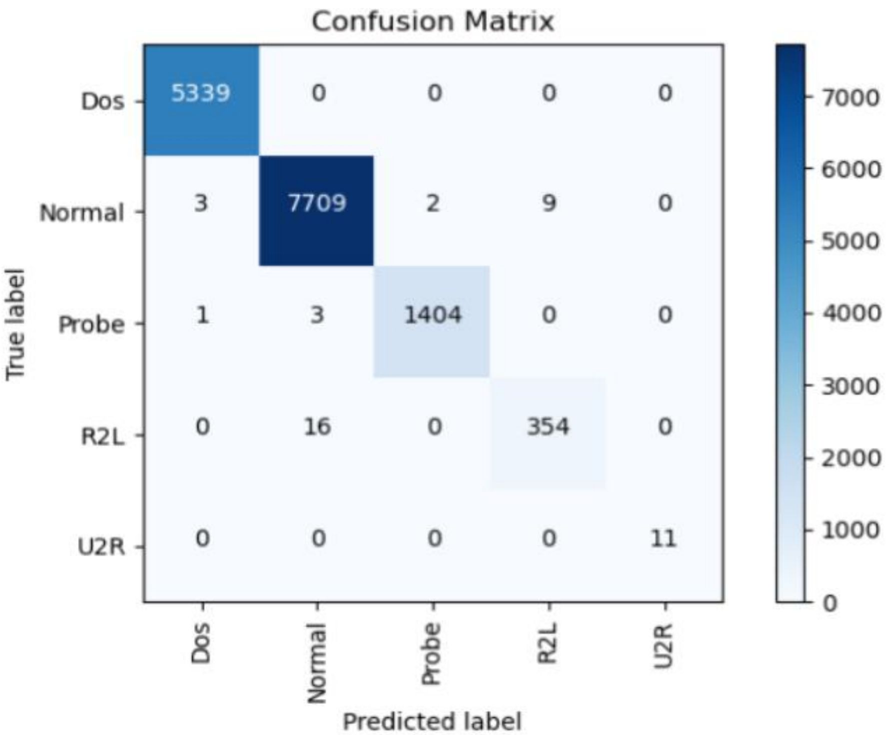


Fig. 6. Confusion matrix for the dataset of NSL-KDD.

TABLE IV. THE COMPARISON OF OUR PRESENTED APPROACH WITH SIMILAR METHODS

Model	ACC	DR	FPR
SVM [41]	69.52	70.23	1.03
RF [42]	69.84	68.79	1.16
AdaBoost [43]	71.98	72.36	0.99
HAST-IDS [44]	93.27	95.85	0.52
CNN-BiLSTM [45]	99.22	98.88	0.43
LuNet [46]	99.14	99.02	0.61
Pelican[48]	99.21	99.13	0.65
Proposed Method	99.81	99.80	0.20

To prove the strong efficiency of the presented approach, the proposed approach of this paper is compared with seven advanced models. These methods are: SVM in [41], RF in [42], AdaBoost in [43], HAST-IDS in [44], CNN-BiLSTM in [45], LuNet in [46] and Pelican in [47]. These methods are trained and tested with a similar data partition and, with the strategy of the cross-validation with k equal to 10. The outcomes of the proposed approach are provided in the best case. Table IV shows the outcomes of the multi-class comparison of the proposed method with seven advanced models. It is clear that the proposed model outperforms the other models on all evaluation criteria. The outcomes display that the presented model betters the detection rate. Also, it maintains a low false positive rate. This point shows the greater effectiveness of the model for network intrusion detection.

C. Discussion

Despite the good performance of the proposed method, there are limitations to the presented work, which are discussed in this section. The main threat to the validity of the proposed method is that only one dataset was used in the evaluation. Therefore, the test results may be different for other datasets that have different types of attack classes. To the best of the authors' knowledge, most publicly available datasets, with the exception of NSL-KDD, are outdated and/or lack raw network traffic data. The proposed approach can extract relevant features from raw network traffic data instead of relying on manual feature selection process. Therefore, it can be easily applied to other datasets. Future work could be to conduct additional experiments by using different datasets such as CSE-CIC-IDS2018 to mitigate this threat.

Another threat to the validity of the proposed method is that the evaluation may appear subjective. Another threat to the proposed work is that only three attack classes are considered. As mentioned, a simple DNN model (ie, a model with a relatively small number of trainable parameters) is trained to detect only certain types of attacks to achieve good accuracy and reasonable prediction time.

V. Conclusion

In the current article, a system of the early detection of the intrusion, which is based on the end-to-end, has been presented for the prevention of real-time network attacks before they cause further detriment to the system of the attack. A classifier basis on CNN is used, to detect the attack. The model is trained as a method of the supervised extraction of the related features by the data of the raw traffic of the network rather than relying on the process of the manual selection of the feature, which is applied to the most relevant methods. The designed experiments have been evaluated in the dataset of NSL-KDD, and the obtained results display that the presented approach improves the overall results (especially the rate of detection and the rate of false positives). The presented approach in the current paper is competitive. It obtains the lower overhead of the computational, which is essential for practical network intrusion detection.

Further analysis has shown that the proposed method has gaps for the betterment of the discerning among the groups by similar features, and the future work will be reviewed here. Furthermore, in the latter task, the researchers can aim to evaluate the proposed approach in the different datasets. Also, the various architectures of the network of the neural can be investigated to check the comparative efficiency in the early detection of the attack.

References

[1] F. Jiang et al., "Deep Learning Based Multi-Channel Intelligent Attack Detection for Data Security," IEEE Trans. Sustain. Comput., vol. 5, no. 2, pp. 204–212, Apr. 2020.

[2] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, “Network intrusion detection system: A systematic study of machine learning and deep learning approaches,” Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, p. e4150, 2021.

[3] Z. Wang, "Deep Learning-Based Intrusion Detection With Adversaries," IEEE Access, vol. 6, pp. 38367–38384, 2018.

[4] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, “Intrusion detection system: A comprehensive review,” Journal of Network and Computer Applications, vol. 36, no. 1, pp. 16–24, 2013.

[5] J. Li, Y. Qu, F. Chao, H. P. H. Shum, E. S. L. Ho, and L. Yang, “Machine learning algorithms for network intrusion detection,” AI in Cybersecurity, pp. 151–179, 2019.

[6] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," Futur. Gener. Comput. Syst., vol. 82, pp. 761–768, May 2018.

[7] Y. Bengio, “Deep learning of representations for unsupervised and transfer learning,” in Proceedings of ICML workshop on unsupervised and transfer learning, JMLR Workshop and Conference Proceedings, 2012, pp. 17–36.

[8] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization.,” ICISSp, vol. 1, pp. 108–116, 2018.

[9] B. Claise, B. Trammell, and P. Aitken, “Specification of the IP flow information export (IPFIX) protocol for the exchange of flow information,” 2013.

[10] A. D. Khairkar, D. D. Kshirsagar, and S. Kumar, “Ontology for detection of web attacks,” in 2013 International Conference on Communication Systems and Network Technologies, IEEE, 2013, pp. 612–615.

[11] Y. Zhang, X. Chen, L. Jin, X. Wang, and D. Guo, “Network intrusion detection: Based on deep hierarchical network and original flow data,” IEEE Access, vol. 7, pp. 37004–37016, 2019.

[12] O. Joldzic, Z. Djuric, and P. Vuletic, “A transparent and scalable anomaly-based DoS detection method,” Computer Networks, vol. 104, pp. 27–42, 2016.

[13] M. Ahmed, A. N. Mahmood, and J. Hu, “A survey of network anomaly detection techniques,” Journal of Network and Computer Applications, vol. 60, pp. 19–31, 2016.

[14] T. F. Ghanem, W. S. Elkilani, and H. M. Abdul-Kader, “A hybrid approach for efficient anomaly detection using metaheuristic methods,” J Adv Res, vol. 6, no. 4, pp. 609–619, 2015.

[15] C. N. Modi, D. R. Patel, A. Patel, and M. Rajarajan, “Integrating signature apriori based network intrusion detection system (NIDS) in cloud computing,” Procedia Technology, vol. 6, pp. 905–912, 2012.

[16] K. Shafi and H. A. Abbass, “An adaptive genetic-based signature learning system for intrusion detection,” Expert Syst Appl, vol. 36, no. 10, pp. 12036–12043, 2009.

[17] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, “An efficient intrusion detection system based on support vector machines and gradually feature removal method,” Expert Syst Appl, vol. 39, no. 1, pp. 424–430, 2012.

[18] Y. Zhang, H. Zhang, X. Zhang, and D. Qi, “Deep learning intrusion detection model based on optimized imbalanced network data,” in 2018 IEEE 18th International Conference on Communication Technology (ICCT), IEEE, 2018, pp. 1128–1132.

[19] B. Yan and G. Han, “Effective feature extraction via stacked sparse autoencoder to improve intrusion detection system,” IEEE Access, vol. 6, pp. 41238–41248, 2018.

[20] Y. Xin et al., “Machine learning and deep learning methods for cybersecurity,” Ieee access, vol. 6, pp. 35365–35381, 2018.

[21] W. Elmasry, A. Akbulut, and A. H. Zaim, “Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic,” Computer Networks, vol. 168, p. 107042, 2020.

[22] A. Aldweesh, A. Derhab, and A. Z. Emam, “Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues,” Knowl Based Syst, vol. 189, p. 105124, 2020.

[23] A. Salari, H. Shakibi, M. Alimohammadi, A. Naghdbishi, and S. Goodarzi, “A machine learning approach to optimize the performance of a combined solar chimney-photovoltaic thermal power plant,” Renew Energy, vol. 212, pp. 717–737, 2023.

[24] Ö. Kasim, “An efficient and robust deep learning based network anomaly detection against distributed denial of service attacks,” Computer Networks, vol. 180, p. 107390, 2020.

[25] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A deep learning approach to network intrusion detection,” IEEE Trans Emerg Top Comput Intell, vol. 2, no. 1, pp. 41–50, 2018.

[26] S. Afzal, B. M. Ziapour, A. Shokri, H. Shakibi, and B. Sobhani, “Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms,” Energy, vol. 282, p. 128446, 2023.

[27] M. Al-Qatf, Y. Lasheng, M. Al-Habib, and K. Al-Sabahi, “Deep learning approach combining sparse autoencoder with SVM for network intrusion detection,” Ieee Access, vol. 6, pp. 52843–52856, 2018.

[28] S. Parampottupadam and A.-N. Moldovann, “Cloud-based real-time network intrusion detection using deep learning,” in 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), IEEE, 2018, pp. 1–8.

[29] Y. Zhang, P. Li, and X. Wang, “Intrusion detection for IoT based on improved genetic algorithm and deep belief network,” IEEE Access, vol. 7, pp. 31711–31722, 2019.

[30] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, “Deep learning approach for intelligent intrusion detection system,” Ieee Access, vol. 7, pp. 41525–41550, 2019.

[31] H. Yang, G. Qin, and L. Ye, “Combined wireless network intrusion detection model based on deep learning,” IEEE Access, vol. 7, pp. 82624–82632, 2019.

[32] M. Elsayed and M. Erol-Kantarci, “Deep reinforcement learning for reducing latency in mission critical services,” in 2018 IEEE Global Communications Conference (GLOBECOM), IEEE, 2018, pp. 1–6.

[33] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, “The expressive power of neural networks: A view from the width,” Adv Neural Inf Process Syst, vol. 30, 2017.

[34] S. Yang and A. Browne, “Neural network ensembles: combining multiple models for enhanced performance using a multistage approach,” Expert Syst, vol. 21, no. 5, pp. 279–288, 2004.

[35] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[36] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.

[37] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.

[38] M. Lin, Q. Chen, and S. Yan, “Network in network. arXiv 2013,” arXiv preprint arXiv:1312.4400, 2013.

[39] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” J Big Data, vol. 6, no. 1, pp. 1–48, 2019.

[40] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[41] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, “Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection,” IEEE access, vol. 6, pp. 33789–33795, 2018.

[42] J. Zhang, M. Zulkernine, and A. Haque, “Random-forests-based network intrusion detection systems,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 5, pp. 649–659, 2008.

[43] W. Hu, J. Gao, Y. Wang, O. Wu, and S. Maybank, “Online adaboost-based parameterized methods for dynamic distributed network intrusion detection,” IEEE Trans Cybern, vol. 44, no. 1, pp. 66–82, 2013.

[44] W. Wang et al., “HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection,” IEEE access, vol. 6, pp. 1792–1806, 2017.

[45] J. Sinha and M. Manollas, “Efficient deep CNN-BiLSTM model for network intrusion detection,” in Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition, 2020, pp. 223–231.

[46] Q. Gao, “Recommended System Optimization in Social Networks based on Cooperative Filter with Deep MVR Algorithm,” 2022.

[47] P. Wu, H. Guo, and N. Moustafa, “Pelican: A deep residual network for network intrusion detection,” in 2020 50th annual IEEE/IFIP international conference on dependable systems and networks workshops (DSN-W), IEEE, 2020, pp. 55–62.

[48] P. Wu, H. Guo, and N. Moustafa, “Pelican: A deep residual network for network intrusion detection,” in 2020 50th annual IEEE/IFIP international conference on dependable systems and networks workshops (DSN-W), IEEE, 2020, pp. 55–62.

The Application of Cognitive Decision-Making Algorithm in Cross-Border e-Commerce Digital Marketing

Xuehui Wang
International School, Zibo Vocational Institute, Zibo, 255000, China

***Abstract*—Extensive global research aims to improve digital marketing profits through pricing decision-making and optimization. A dual word-of-mouth diffusion pricing model is developed for cross-border e-commerce, addressing word-of-mouth accumulation and information diffusion effects. The traditional artificial bee colony algorithm is optimized with security domain search and information diffusion profiles, enhancing global search capabilities. Performance tests reveal that word-of-mouth scale significantly influences cross-border e-commerce profits, increasing with scale coefficient, consumer conversions, and optimal profits. The proposed algorithms demonstrate high efficiency and convergence rates, surpassing common iterations and benefits in the clothing pricing problem. The comprehensive imitation effect is -0.14, and the word-of-mouth scale effect is 1.34. Pre-sale and sales prices for clothing are set at 347.49 and 641.393, respectively. Similarly, in pricing cross-border e-commerce electronic products, the algorithm achieves optimal profits after 230 iterations, surpassing other algorithms. Overall, the proposed model exhibits superior computational performance in cross-border e-commerce pricing decision-making compared to conventional approaches.**

***Keywords*—Cross-border e-commerce; decision-making; pricing issues; optimization algorithms; ABO**

I. INTRODUCTION

With the year-on-year increase in Internet users, e-commerce, especially cross-border e-commerce (CBEC), has developed rapidly in the context of the normalisation of the new crown epidemic. The spread of the epidemic and the preventive and control measures of various countries have had a serious impact on international trade activities. The existing supply chain and industrial chain have been obstructed. The game between China and the United States and the changes in the international political situation have made the international market economically volatile. However, the development of e-commerce has brought prosperity to China's CBEC business, which has become an important force in stabilising foreign trade and played an important role in stabilising global supply [1-3]. CBEC has become increasingly competitive, with international companies such as Amazon and Ebay, as well as domestic companies such as Jingdong and Tmall, entering the field [4-5]. In CBEC digital marketing, the application of cognitive decision-making (CDM) algorithms is of great value. Li et al. designed a hierarchical product classification and retrieval system suitable for CBEC shopping websites based on the analysis of image retrieval algorithms. The classification decision layer was used to determine the category of product images, and then the corresponding product image features were accurately retrieved. The recommendation results were highly accurate [6]. This algorithm is based on modeling the decision-making of individuals and exploring the theory and methods of how individuals make decisions under uncertainty. The study aims to explore the application of CDM algorithms in CBEC digital marketing to improve marketing effectiveness and economic efficiency. The study includes an introduction to CBEC digital marketing and CDM algorithms, a study of the pricing strategy problem in CBEC digital marketing, a test and analysis of the performance of the models and algorithms, as well as a summary and exposition of the study. The main contribution of the research is the application of CDM algorithms to CBEC digital marketing, which provides a new perspective in the field of international trade to improve marketing effectiveness and economic efficiency.

II. RELATED WORKS

CBEC and digital marketing have been hot topics in the economic field in recent years, and a large number of scholars have conducted research on them. Setkute et al. conducted research on digital marketing inapplicability in B2B small and medium-sized enterprises, and used qualitative research methods to investigate background factors of small B2B companies. They analyzed the obstacles that affect digital marketing practices, and the results showed that the “one size fits all” digital marketing mindset is not suitable for B2B small and medium-sized enterprises [7].

Yan et al. analyzed the impact of performance management systems on employee productivity in Chinese CBEC enterprises by using quantitative methods. Descriptive statistics were used as a data analysis tool to analyze the statistical data of 400 employees of the surveyed e-commerce enterprises using percentages and frequencies [8]. Li et al. conducted research on the impact of environmental regulation on CBEC exports of agricultural products. Based on the certification process of agricultural products under the influence of environmental regulation, they established an agricultural products output equation under the influence of environmental regulation, and analyzed how environmental regulation intensity affects the quality of agricultural products exported through CBEC and the competitiveness of agricultural products exporting enterprises [9]. CDM algorithms are based on modeling the decision-making process of individuals and exploring the theories and methods of how individuals make decisions under uncertainty. They are widely used in various fields. Ezaleden et al. used the NPSO algorithm to learn the importance of relationship types between concepts to complete a simulated recommendation system based on the highest ranking for dynamic learners. They studied the CLM and ECLM concept models, and the simulation results showed that ECLM performed better than other existing methods, with a mean reciprocity rate value of 0.780 [10].

Cao et al., to achieve intelligent recognition of surface Electromyography (sEMG) gesture signals in human-computer interaction, proposed a sEMG gesture recognition intelligent model combined with feature extraction, genetic algorithm (GA) and SVM model, and proposed adaptive mutation particle swarm optimization (AMPSO) algorithm to optimize SVM parameters. Research outcomes denoted that AMPSO-SVM could effectively recognize low-frequency sEMG signals of different gestures, with good performance [11]. Stojanovic Blaza et al. applied intelligent optimization algorithms to stability control of multi machine power systems. The stability of this method in system dynamic stability control has been demonstrated through comparative experiments of simulation results [12]. Jarndal et al. applied PSO and GA intelligence optimization algorithms to the efficient electrothermal large signal GaN HEMT modeling. Experiments have shown that the model also exhibited accurate simulation of nonlinear power amplifiers, with excellent computational speed and convergence [13]. Song et al. coupled the temperature and structure of the braking system using finite element method, and used GA parameter optimization and sensitivity analysis. Experiments have shown that this method can optimize the thermal stress and deformation problems of fan openings in high-temperature environments [14].

In summary, although researchers have conducted extensive research on various aspects of CBEC and digital marketing, research in digital marketing decision-making is still very scarce, which is highly related to CBEC profits. Using CDM algorithms to conduct due research on it has high potential application value.

III. Design of CBEC Digital Marketing Pricing Decision Model

With digital marketing as the support, the digital marketing manages the production, logistics, distribution, publicity and a series of enterprise marketing activities that run through the product cycle online and offline. The marketing channel tends to be flat, and the direct communication between enterprises and consumers will cost business operations. Product pricing is the basic point of digital marketing and the main factor for scholars to analyze enterprise marketing decisions. At present, the research on product pricing decisions focuses on the two-stage sales model of “pre-sale+sales”, and the research will also study the digital marketing pricing under this sales model.

A. Design of Product Pricing Model for the Diffusion Benefits of Cross-border e-commerce Digital Marketing

The research on product pricing focuses on the game and equilibrium between consumer behaviour and enterprises' pricing. In the two-stage sales model of “pre-sale+sales”, it is mainly the impact of enterprises' pricing diffusion on the demand of the later market. The pattern of this diffusion event is shown in Fig. 1.

TABLE I. VARIABLE SYMBOLS

Symbol	Meaning; implication; sense; import; meanings
c	The production cost for each product sold
p_1	Pre-sale price
p_2	Sales period price
N	Number of consumers receiving product information (i =1,2)
u	Utility values obtained per unit of product
θ	Strategic consumers in the pre-sale period of the proportion (0< θ <1)
Δ	Utility values obtained by product diffusion
Q	Order quantity of products
m	Regret utility
δ	The possibility that consumers will buy a product during the normal sales period
D_1	Product demand during the pre-sale period
D_2	Crystal production demand during the sales period
Π_1	Product profit during the pre-sale period
Π_2	Crystalline production profit during the sales period.
β_1	The effect coefficient of the innovation effect, ($\beta_1 > 0$)
β_2	The influence coefficient of the imitation effect, ($\beta_2 < 0$)
β_w	Integrated imitation effect coefficient
η	Scale effect coefficient
$F(.)$	The probability of consumers buying a product during the pre-sale period.
N_{D1}	The number of consumers attracted by the product diffusion effect
g_r	Out of stock cost of unit product due to out of stock
$k(.)$	The impact of crystal production diffusion on retailers' normal sales period demand
φ	The percentage of negative word of mouth

In the pre-sale stage of products, CBEC publishes information such as pre-sale and normal sales prices, and quantities on e-commerce platforms. To attract consumers to purchase, CBEC sets pre-sale prices lower than normal sales prices. However, due to the limitations of information dissemination on e-commerce platforms, it is not possible for all consumers to receive pre-sale related information. Only some well-informed consumers can successfully receive product pre-sale information, while consumers who have not received product pre-sale information are classified as message blocking consumers. Informed consumers judge the effectiveness of products based on product information, measure prices and demand, and make decisions. Consumers who purchase products publish their usage experiences on the internet, and over time, their word-of-mouth (WoM) accumulates or decreases, leading to the diffusion of product information. This diffusion effect affects the number of consumers attracted during the normal sales phase [15-16]. The diffusion model of Base is shown in Fig. 2.

The process of converting potential purchasing groups is influenced by two aspects: innovation and imitation effects. Innovation effect mainly refers to the influence of external factors such as advertising, marketing, and price on shopping behavior, while imitation effect is influenced by WoM to cause potential groups to follow the buying behavior. When product diffusion events occur, CBEC enterprises make decisions on product supply based on factors such as reputation and the number of pre-sale consumers in the product diffusion effect, to minimize cross-border logistics storage costs and achieve max profits. Because the product diffusion effect will affect the actual demand in the pre-sale+sales period, the optimal pricing decisions for enterprises to obtain max profits are divided into pre-sale and sales price decisions. The variable symbols used in the decision model are displayed in Table I.

Under the diffusion effect, the optimal selling price decision of a product can be expressed as Eq. (1).

$$p_2^* = \begin{cases} \dfrac{\lambda + \theta\lambda\bar{F}(p_2) + N_{D1}}{\theta\lambda f(p_2)}, & N_{D1} \le Q - 2\lambda \\ \dfrac{Q - (1-\theta)\lambda - \theta\lambda F(p_2)}{\theta\lambda f(p_2)}, & N_{D1} > Q - 2\lambda \end{cases} \tag{1}$$

In Eq. (1), $\bar{F}(p_2) = 1 - F(p_2)$ denotes the proportion of consumers who purchase products during the sales period. It should be noted that there is a situation where $\min(Q - (1-\theta)\lambda - \theta\lambda F(p_2), \theta\lambda\bar{F}(p_2) + \lambda + N_{D1})$ is taken. When supply exceeds demand, i.e. $N_{D1} \le Q - 2\lambda$, the sales profit of CBEC is as expressed in Eq. (2).

$$\begin{aligned} E(\Pi_2) = {} & p_2(\lambda + \theta\lambda\bar{F}(p_2) + N_{D1}) \\ & -c(Q - (1-\theta)\lambda - \theta\lambda f(p_2)) \end{aligned} \tag{2}$$

If the second-order derivative is less than 0, then the profit function has an optimal sales period price. When demand exceeds supply, there is $N_{D1} > Q - 2\lambda$. Since $\bar{F}(p_2) = 1 - F(p_2)$ is a decreasing function of p_2, the optimal sales period price relationship is shown in Eq. (3).

$$\begin{aligned} E(\Pi_2') = {} & (p_2 - c)(Q - (1-\theta)\lambda - \theta\lambda F(p_2)) \\ & -g_r(2\lambda + N_{D1} - Q) \end{aligned} \tag{3}$$

From Eq. (3), it is easy to obtain that its second-order derivative is less than 0, and the function has the optimal sales period price. From the above analysis, the product price decision during the sales period of CBEC mainly depends on the impact of diffusion effects on consumer demand. When the impact is small, there is still surplus in the product, and the sales price will increase with the increase of diffusion. On the contrary, if the product supply is insufficient, the setting of sales prices can temporarily ignore the diffusion effect of the product during the pre-sale period. The expression for the optimal pre-sale price of a product under the diffusion effect is shown in Eq. (4).

$$\begin{aligned} p_1^* &= u - (u - p_2^* + \Delta)\delta + m(1-\delta) \\ &= (1-\delta)(u+m) + (p_2^* - \Delta) \end{aligned} \tag{4}$$

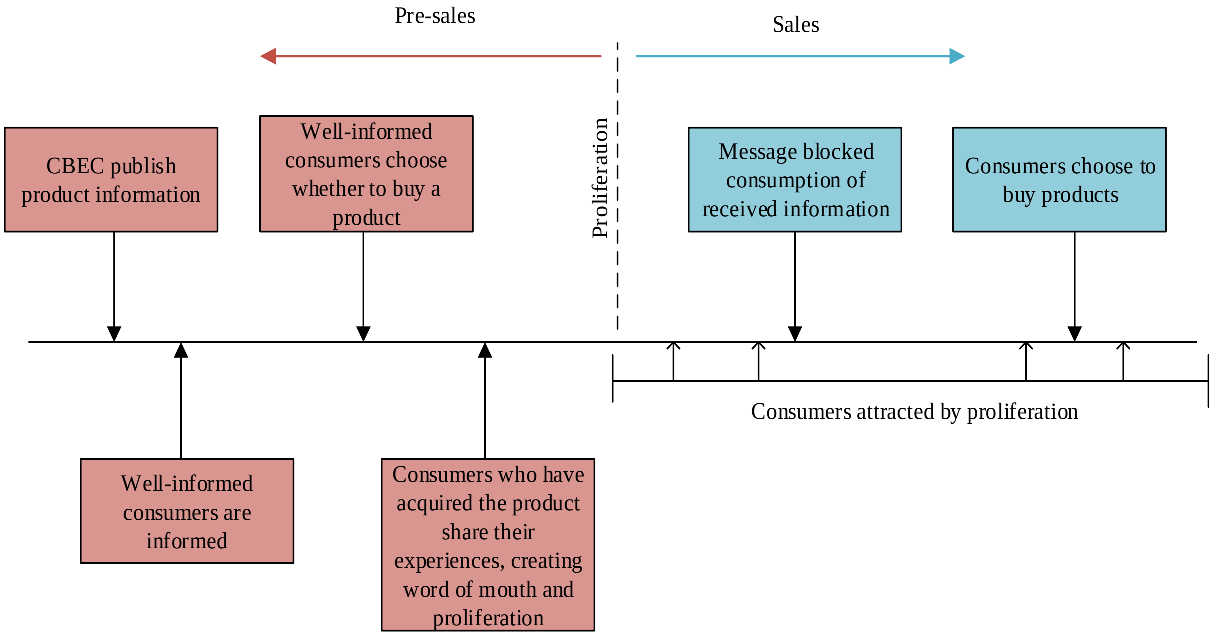
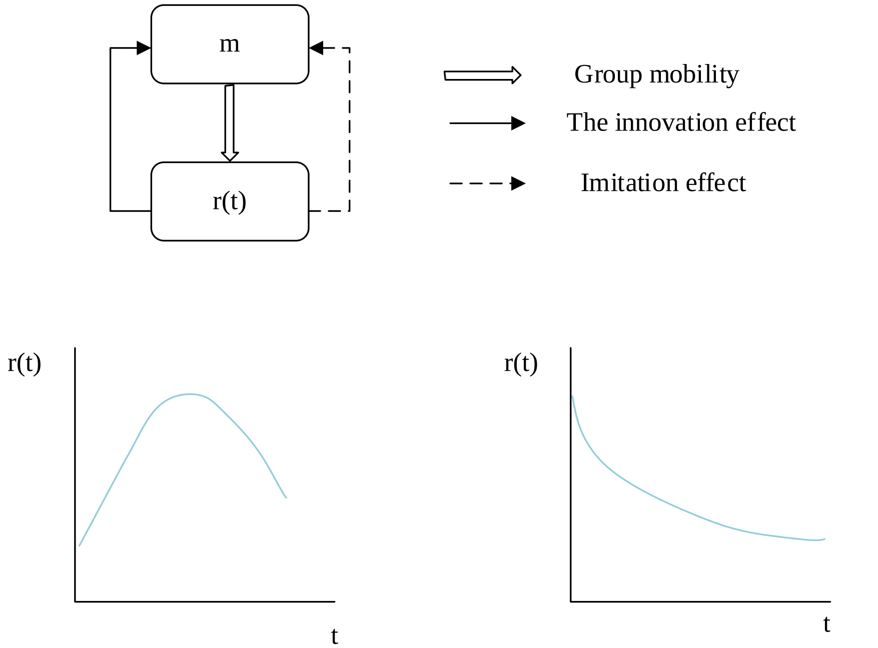

Pre-sales
Sales
CBEC publish product information
Well-informed consumers choose whether to buy a product
Proliferation
Message blocked consumption of received information
Consumers choose to buy products
Consumers attracted by proliferation
Well-informed consumers are informed
Consumers who have acquired the product share their experiences, creating word of mouth and proliferation

Fig. 1. Event diffusion model.



(a) The imitation factor is greater than the innovation factor

(b) The innovation factor is greater than the imitation factor

Fig. 2. Base diffusion model.

The pre-sale pricing of CBEC is influenced by the sales period price. As the diffusion effect increases on consumer demand, when supply exceeds demand, the sales price increases accordingly. The pre-sale price is higher, but when supply is less than demand, it has little impact on the pre-sale price. The scale effect of WoM refers to the impact of WoM generated by the sale of a product on the purchasing intention of potential consumers, increasing (or decreasing) the inflow of intended consumers, and the potential consumers who form the scale effect of WoM are ηN_2. Consumers who purchase products will be divided into positive and negative WoM groups, which is known as the WoM ratio effect. A product pricing model that combines the scale and proportion of WoM to form a dual WoM diffusion effect. The impact of product diffusion on the normal sales period demand of CBEC in the model is shown in Eq. (5).

$$\begin{aligned} k(D_1) = \beta_1 + \beta_2 + \frac{\varphi((1-\theta)N_1 + \theta\lambda F(p_2))}{\eta N_2} \\ -\beta_2 - \frac{\varphi((1-\varphi)(1-\theta)N_1 + \theta N_1 F(p_2))}{\eta N_2} \end{aligned} \tag{5}$$

In Eq. (5), β_1 and β_2 denote the coefficients of innovation and imitation effects, respectively. The coefficient of innovation effect refers to the coefficient of external factors other than WoM that affect consumer shopping behavior. The coefficient of imitation effect stands for the coefficient of shopping behavior affected by WoM transmission, which is divided into positive WoM imitation effect coefficient β_{2+} and negative WoM imitation effect coefficient β_{2-}. The comprehensive imitation effect coefficient of the two is denoted in Eq. (6).

$$\beta_w = (1-\varphi)\beta_{2+} - \varphi\beta_{2-} \tag{6}$$

Under the dual WoM diffusion effect, the impact of WoM diffusion on customer demand during CBEC sales is shown in Eq. (7).

$$k(D_1) = \beta_1 + \beta_w \frac{(1-\theta+\theta F(p_2))}{\eta} \tag{7}$$

The dual effects of WoM and diffusion effects jointly constitute the consideration factors for CBEC pricing decisions, considering multiple supply and demand relationships. Under the influence of the dual WoM diffusion effect, the sales price can be calculated as expressed in Eq. (8).

$$p_2^{**} = \begin{cases} \dfrac{(1+\theta+\beta_1) + \dfrac{\beta_w}{\eta}(1-\theta)(1+F(p_2)) + c\theta f(p_2)}{\theta f(p_2) - \dfrac{\beta_w}{\eta}\theta f(p_2)}, \\ \dfrac{Q}{\lambda} - 2 \geq \beta_1 + \dfrac{\beta_w}{\eta}(1-\theta+\theta F(p_2)) \\ \quad \dfrac{Q-(1-\theta)\lambda - \theta\lambda F(p_2)}{\theta\lambda f(p_2)} + c + g_r(1-\dfrac{\beta_w}{\eta}), \\ \quad \dfrac{Q}{\lambda} - 2 < \beta_1 + \dfrac{\beta_w}{\eta}(1-\theta+\theta F(p_2)) \end{cases} \tag{8}$$

According to Eq. (8), when supply exceeds demand, p_2^{**} is positively correlated with β_1, and is influenced by β_w and η. That is, the higher the external influence of WoM on consumers, the greater the diffusion effect. At this point, a high pricing decision can be chosen. When supply and demand are not met, the sales price is highly correlated with the existing reputation of the product. Therefore, the premise for the optimal pricing of CBEC is as Eq. (9).

$$p > c + g_r(1 - \frac{\beta_w}{\eta}) \tag{9}$$

Under the dual WoM diffusion effect, the pre-sale price of the product is set as shown in Eq. (10).

$$p_1^{**} = u - (u - p_2^{**} + \Delta)\delta + m(1-\delta) \tag{10}$$

In summary, the pricing of products during the pre-sale period of CBEC is influenced by the dual WoM diffusion effect. When the supply is sufficient, the increase in innovation effect has a significant impact on potential consumers, and the pre-sale price is higher. The product diffusion effect is influenced by comprehensive WoM, which also affects product pricing. When supply does not meet demand, pricing is only related to WoM.

B. *CD-ABC Algorithm Design Based on Security Domain Search Strategy and Information Diffusion Statistical Model*

The products and markets in CBEC activities are relatively complex, requiring a deep understanding of consumer behavior and preferences to make better decisions. There is a high demand for the global optimization ability of CDM algorithms. Therefore, an artificial bee colony (ABC) algorithm with strong global optimization ability, high solving accuracy, and few control parameters is selected to solve the pricing decision problem. The ABC algorithm is derived from human observation of the information exchange process during bee foraging. Karaboga proposed the ABC algorithm based on this process, dividing individuals searching for the target space into three roles: picking bees, observing bees, and reconnaissance bees. These three roles switch to each other as needed. Picking bees search for new foraging locations based on known information and share it with observing bees. Reconnaissance bees are responsible for randomly searching for new honey sources near the hive. In the ABC algorithm, the dimension of the solution to the optimization problem is D dimension, and one solution of the problem is the coordinate corresponding to a honey. The amount of honey is the fitness of the solution, and the number of honey and bees collected or observed is equal, set as SN. The process of searching for the next honey position is as shown in Eq. (11) when the bees reach one honey position [17-20].

$$x_{id} = x_{id} + \varphi_{id}(x_{id} - x_{kd}) \tag{11}$$

In Eq. (11), $i = 1, 2, \cdots, SN$, $d = 1, 2, \cdots, D$, and φ_{id} are random numbers of [-1.1], and $k \neq i$. After picking bees to find new honey, it compares the new honey position with the original honey position, and retains the position with the highest amount of honey. If the new honey quantity is lower than the old honey, the information is transmitted to the observing bee [21]. It observes the bees and calculates the probability of the next honey occurrence position based on Eq. (12).

$$p_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j} \tag{12}$$

Picking and observing bees will traverse the honey in the current domain. If the fitness value of the honey does not improve before reaching the limited number of iterations, the honey will be discarded. At the same time, the bees in this position will be transformed into reconnaissance bees, and the fit_i in the equation represents the fitness value of the solution. After the process is completed, the reconnaissance bee will search for the next honey in the solution space, as shown in Eq. (13).

$$x_{id} = x_d^{\min} + r(x_d^{\max} - x_d^{\min}) \tag{13}$$

In Eq. (13), $x_d^{\max}$ and $x_d^{\min}$ denote the upper and lower limits of the D-dimensional solution space, and r means a random number in the interval [0,1]. The reconnaissance bee searches in the solution space by randomly selecting a number in the D-dimensional solution space to obtain a location. The ABC algorithm flow is shown in Fig. 3.

The ABC algorithm starts by randomly initializing the bee colony, calculating the amount of honey at the location where the bees are collected, and then the bees start searching for the next honey to update the honey location. At the same time, the bees select the honey location for observation. To ensure that bees are not attacked (interfered) by other groups when searching for new honey sources, a security domain search strategy is studied, as shown in Eq. (14).

$$v_{id} = x_{id} + r(x_{id} - x_{kd}) + r(SP_d - x_{id}) \tag{14}$$

In Eq. (14), SP_d indicates the safe location of the d honey source. Obviously, v_{id} is randomly guided from multiple directions and approaches the safe location. The safe location in the group is expressed in Fig. 4.

In a bee colony, the area where the proportion of bees to the population is greater than μ is considered a safe area. The triangle in the figure represents bees, and the circle centered around the safe position C_d is considered a safe area. The maximum safe distance of the $C_j = SP = (SP_1, SP_2, \cdots, SP_D)$ safe area is the radius MSD of SP. In the observing bee search stage, it conducts probability selection according to the honey source information transmitted by the picking bee to find the optimal location for mining. However, this method is uncertain and blind, and cannot ensure that good honey sources are more attractive to observing bees than poor honey sources. The statistical model of information diffusion is introduced to adjust the selection strategy of the picking bee, as shown in Eq. (15).

$$p(x_i) = \frac{1}{(N-1)\sqrt{2\pi}} e^{\frac{-(fit(x_i) - fit_{\max})^2}{2h^2}} \tag{15}$$

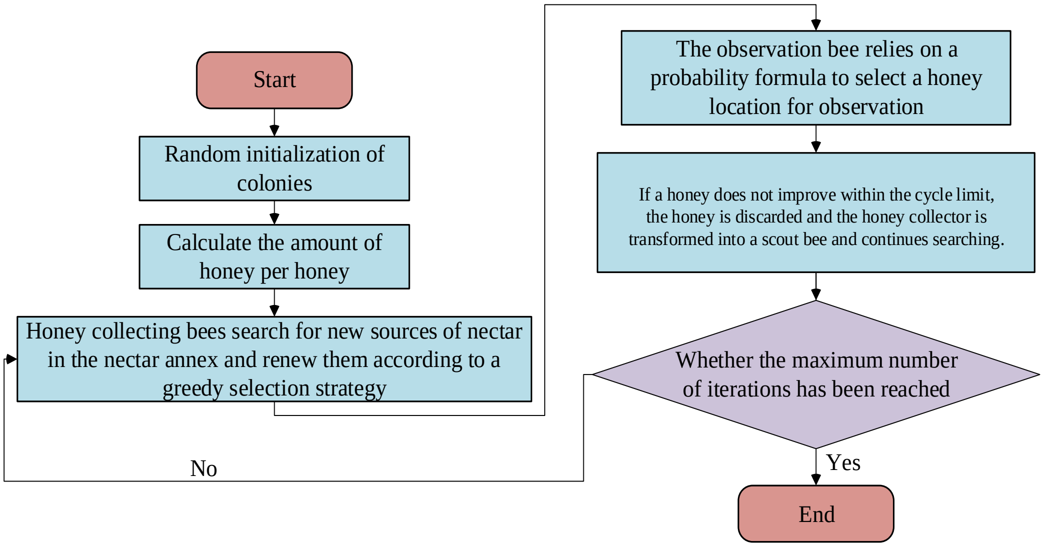

Start
Random initialization of colonies
Calculate the amount of honey per honey
Honey collecting bees search for new sources of nectar in the nectar annex and renew them according to a greedy selection strategy
The observation bee relies on a probability formula to select a honey location for observation
If a honey does not improve within the cycle limit, the honey is discarded and the honey collector is transformed into a scout bee and continues searching.
Whether the maximum number of iterations has been reached
No
Yes
End

Fig. 3. ABC algorithm flowchart.

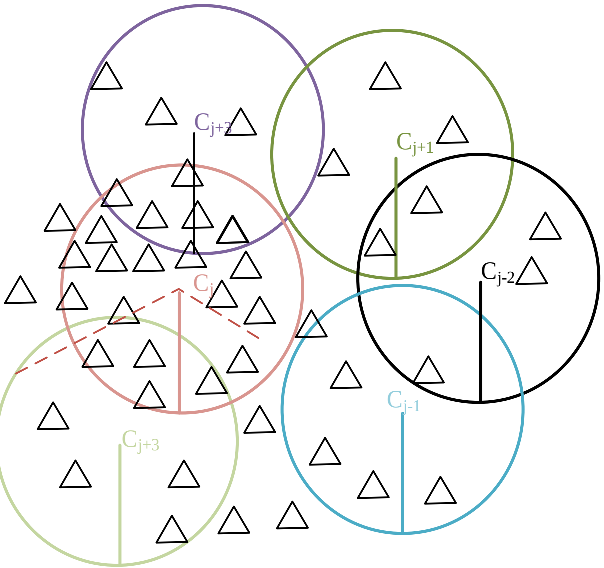


Fig. 4. Safe location in a group.

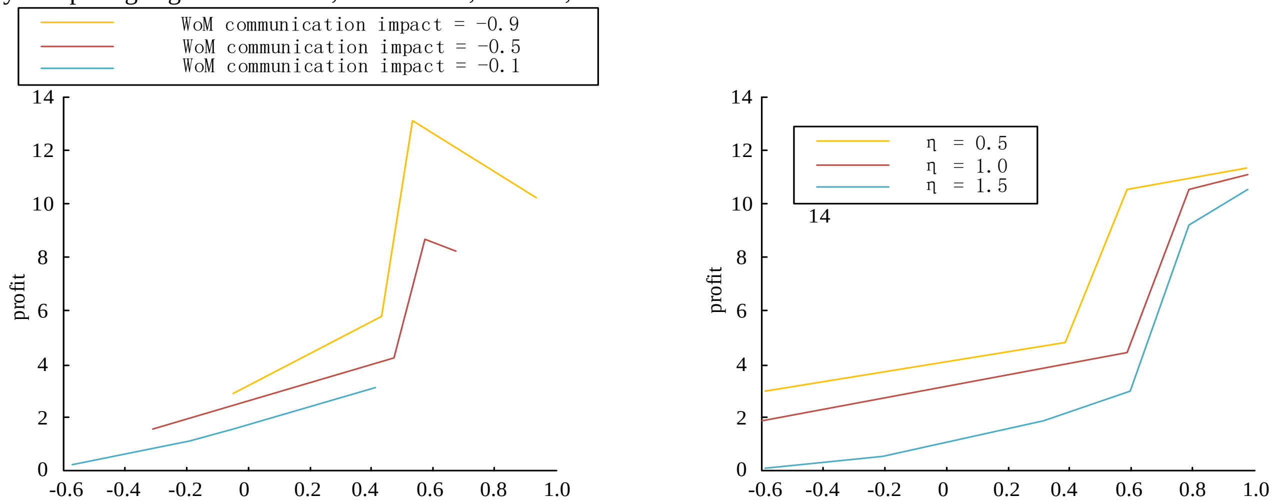
In Eq. (15), $p(x_i)$ refers to the probability of observing bees selecting honey source i, and h stands for the information diffusion coefficient.

IV. Performance Test of CBEC Digital Marketing Pricing Decision Model

Based on research on CBEC products, a clothing product of a certain CBEC enterprise was selected as an example for model performance testing. First, the relationship between innovation effect coefficient and pre-sale pricing under different WoM propagation of product pricing model, and the relationship between comprehensive imitation coefficient and CBEC pricing under different word of mouth scale were analyzed. The impact of product diffusion effect and comprehensive imitation effect on CBEC profits was investigated. The results are shown in Fig. 5.

As shown in Fig. 5, Fig. 5(a) and Fig. 5(b) show the impact of product diffusion effect and comprehensive imitation coefficient on CBEC profits, respectively. Fig. 5(a) shows that the impact of WoM communication has increased, and the profits obtained by CBEC have increased. At the same time, as the product diffusion effect increased, the profits obtained by CBEC continued to increase. Observing the diffusion effect coefficient of products, as the effect coefficient increased, supply exceeded demand, sales decision prices increased, and pre-sale prices were higher. At this time, CBEC profits increased. When supply did not meet demand, the increase in diffusion effect only increased the gap in goods, leading to a decrease in CBEC profits. Fig. 5(b) shows the impact of WoM scale effect on CBEC profits under the dual WoM diffusion effect. As the WoM scale coefficient increased, the amount of potential consumer conversions increased, and the optimal profits obtained by CBEC increased. As the comprehensive imitation coefficient increased, the diffusion effect of products increased and CBEC profits increased due to the influence of WoM scale effect. The rationality of the model was verified. Based on this, the CD-ABC algorithm's wide area search capability and resource utilization capability were tested to verify the excellent performance of the algorithm. The C-ABC algorithm (secure neighborhood search ABC algorithm), D-ABC algorithm (information diffusion ABC algorithm) and ABC algorithm were used as comparison objects. Twenty tests were conducted on 22 benchmark functions (D=30). The population setting was 30, $\lim it = 200$, and the maximum number of iterations was 100000. The test results are shown in Fig. 6.

By comparing Fig. 6 as a whole, the D-ABC, C-ABC, and CD-ABC algorithms generally outperformed the ABC algorithm in most functions. In most functions, the convergence degree of the D-ABC and C-ABC algorithms was better than that of ABC, indicating that compared to traditional ABC algorithms, the two improved methods proposed in the study had good performance optimization. CD-ABC algorithm had better performance than C-ABC and D-ABC algorithms in most functions, and its rate of convergence was significantly faster than D-ABC and C-ABC, indicating the effectiveness of the improved CD-ABC algorithm. To further illustrate the advantages of information diffusion probability over ABC probability, a test was conducted on the impact of bees on the honey source search probability for the same initial population. The results are shown in Fig. 7.



(a)The impact of product diffusion effects on the profitability of CBEC

(b)The impact of the combined imitation factor on CBEC profits

Fig. 5. The impact of product diffusion effects and combined imitation effects on the profitability of CBEC.

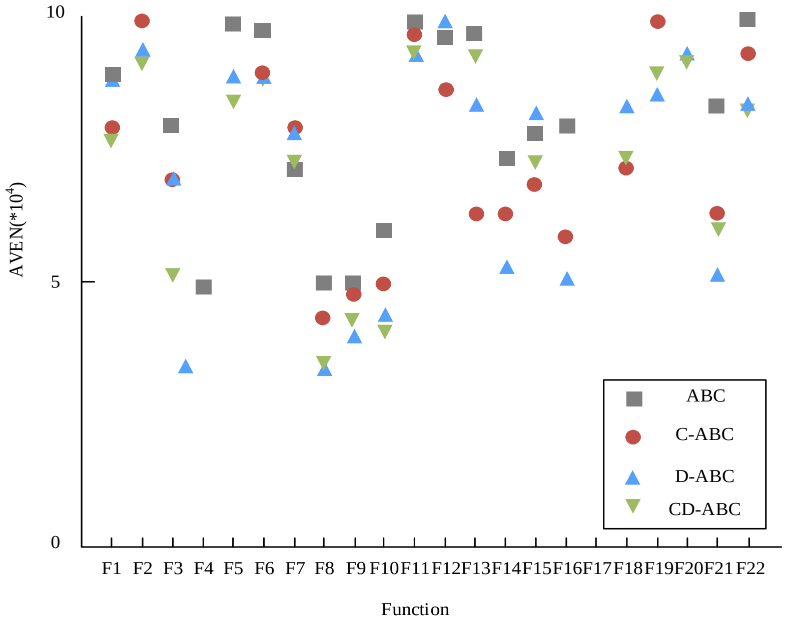
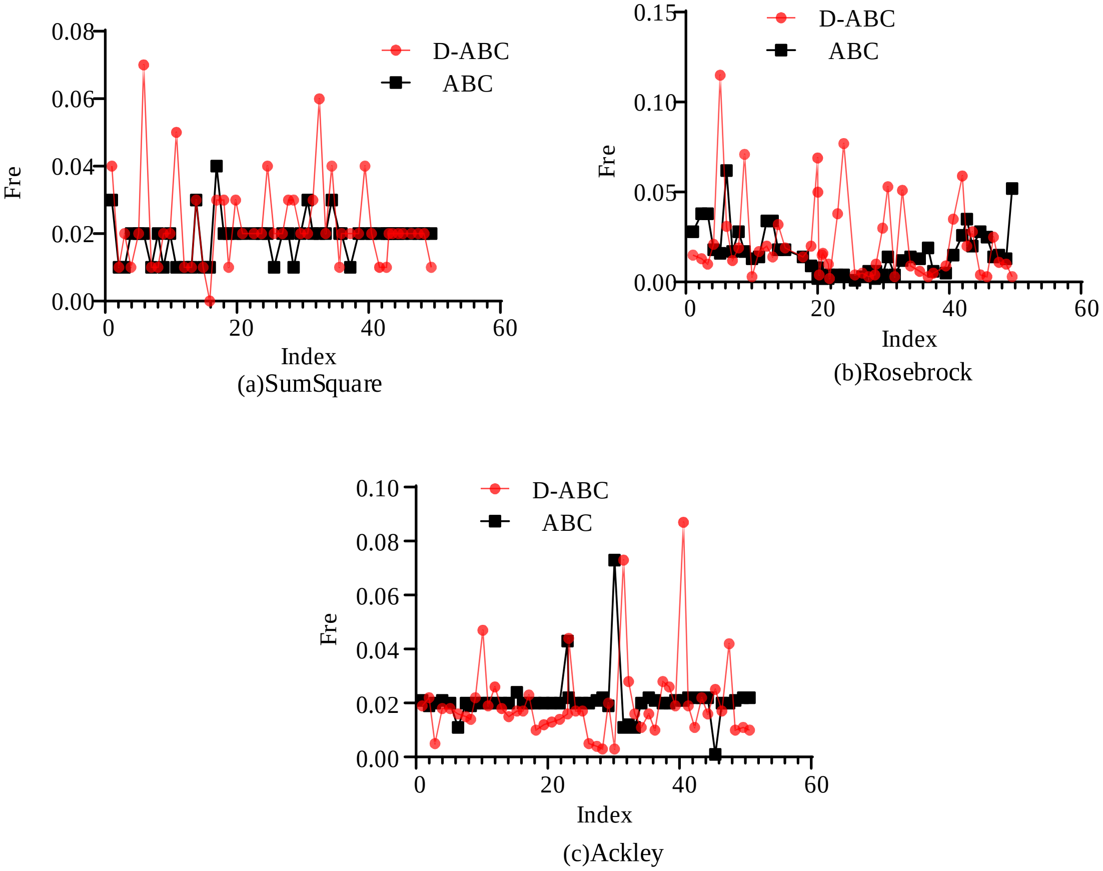


Fig. 6. Algorithm average AVEN comparison.



(a)SumSquare

(b)Rosebrock

(c)Ackley

Fig. 7. Individual evolution frequency test.

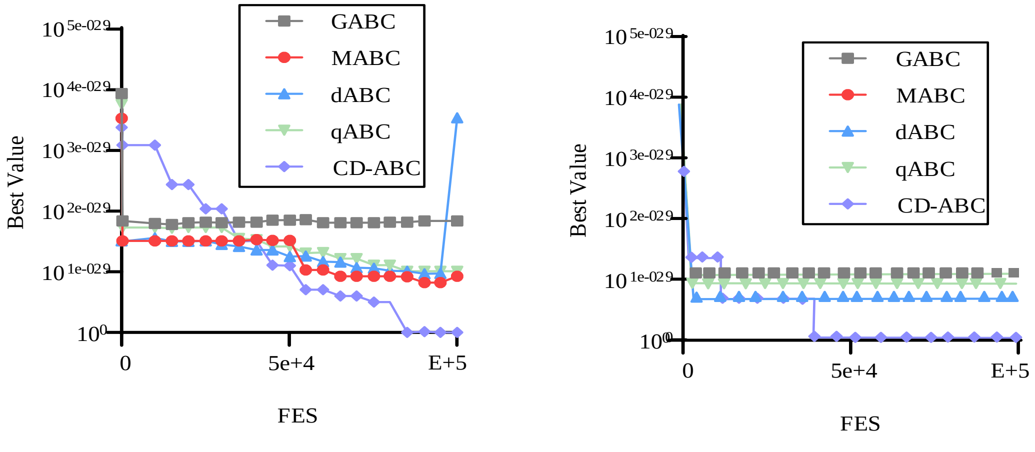
As shown in Fig. 7, representative unimodal functions SumSquare, uncertainty function Roserock, and multimodal function Ackley were selected for experiments to test the ABC algorithm's concept. Comparing Fig. 7 as a whole, there was no significant difference in the individual evolution frequency of the ABC algorithm, indicating that it was unable to detect the superiority of honey sources. However, in D-ABC, some individuals exhibited higher evolution frequencies, indicating that the information diffusion probability could detect the superiority of honey sources and convey it to the observation bees, so that the excellent and beautiful honey sources received more attention. This indicated that the information diffusion probability could detect the value of hidden honey sources in ABC, improving ABC's deep mining ability. Based on the above test results, the CD-ABC algorithm was compared and tested with other ABC variant algorithms with excellent improvement effects. The parameter settings and algorithm are displayed in Table II.

to conduct a fair comparison test on the above improved algorithms, the same parameter settings were used for the experiment. The results of the algorithm rate of convergence comparison test are shown in Fig. 8.

When D is set to 30, the CD-ABC algorithm outperformed GABC, MABC, qABC, and dABC in 14 functions. Fig. 8 shows the convergence curves of the above algorithms on some functions. CD-ABC showed high efficiency and Rate of convergence in most functions. Based on the above verification of the efficient performance of CD-ABC, its performance in solving practical pricing problems was tested. The product pricing model with dual WoM diffusion effect proposed in the study was used for solution analysis, and compared with particle swarm optimization (PSO), whale optimization algorithm (WOA), GA, ABC and WOA-GA algorithms. The experimental object was the pricing of clothing products in CBEC, as shown in Fig. 9.

TABLE II. ALGORITHM COMPARISON TEST PARAMETER SETTINGS

Improved algorithms	Parameter settings		
	N	limit	C/P/r/u
GABC	50	200	1.5
MABC	50	200	0.7
dABC	50	200	/
qABC	50	200	1.5
CD-ABC	50	200	0.3



(a)Sphere

(a)Elliptic

Fig. 8. Comparison test results of algorithm rate of convergence.

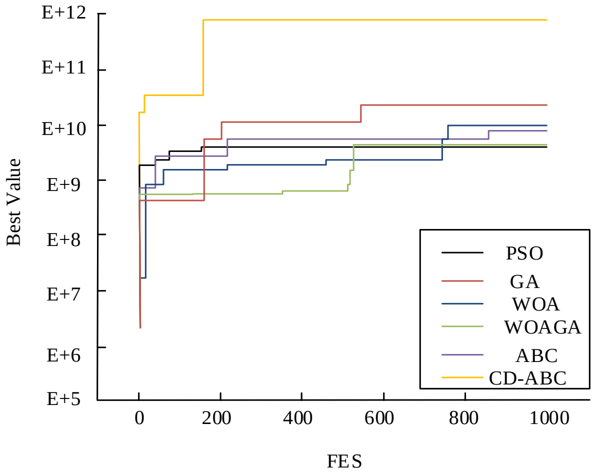


Fig. 9. Iterative curves for solving clothing pricing problems using different algorithms.

Fig. 9 shows the iteration curve of each algorithm to solve the CBEC clothing pricing problem under the dual WoM diffusion effect model. The CD-ABC algorithm obtained the optimal pricing decision, and its rate of convergence was fast. After 200 iterations, the convergence was completed, and the number of iterations was far lower than the common decision solving algorithm. And the pricing optimization decision obtained by the algorithm resulted in CBEC gained much higher profits than other algorithms, ultimately benefiting 7.95e+11. The comprehensive imitation effect calculated by the CD-ABC algorithm in 200 iterations was -0.14, the reputation scale effect was 1.34, the pre-sale price was finally set at 347.49, and the sales price was 641.393, which indicated that the pricing obtained by the CD-ABC algorithm was closer to life and could provide a reference for enterprises to make pricing decisions. To demonstrate the applicability of the CD-ABC algorithm in solving pricing decision-making problems in various application scenarios, a CBEC enterprise's electronic product was selected as a case for comparative testing. The test results are shown in Fig. 10.

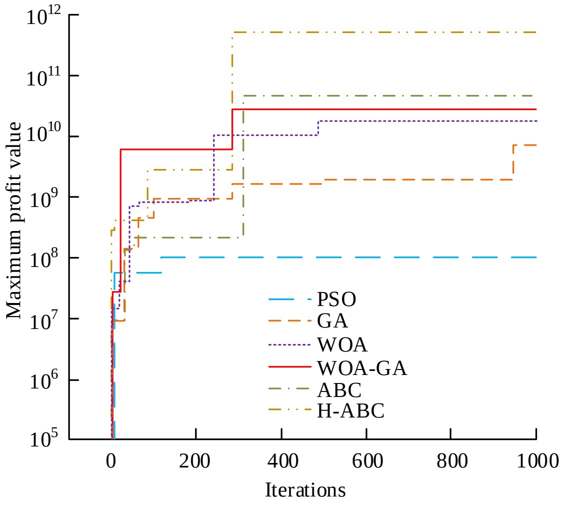


Fig. 10. Iterative curve for solving the pricing problem of cross border e-commerce electronic products.

Fig. 10 shows the iteration curve of each algorithm to solve the pricing problem of CBEC electronic products under the dual WoM diffusion product pricing model. The CD-ABC algorithm obtained the maximum profit. After 230 iterations, the optimal profit decision reached 5.31e+11, which was far higher than the results of other algorithms. However, its rate of convergence was worse than that of GA, and its global optimization ability was weak. However, overall, it still had high performance in solving CBEC pricing problems, which could provide a certain basis for CBEC pricing decisions.

V. Conclusion

To optimize CBEC marketing strategies and improve competitiveness, the research focused on the pricing throughout the whole e-commerce marketing, built a dual WoM diffusion product pricing model suitable for CBEC product pricing, and optimized the traditional ABC algorithm to enable it to have excellent performance in solving the dual WoM diffusion product pricing model strategy. The performance test outcomes of the model indicated that the impact of WoM communication increased, and the profits obtained by CBEC increased. At the same time, as the product diffusion effect increased, the profits obtained by CBEC continued to increase. Under the dual WoM diffusion effect, the impact of WoM scale effect on CBEC profits increased with the coefficient of WoM scale, the number of potential consumer conversions, and the optimal profits obtained by CBEC. Using the C-ABC, D-ABC and ABC algorithms as comparison objects, the test results of 22 benchmark functions showed that the D-ABC, C-ABC and CD-ABC algorithms generally outperformed ABC algorithm in most functions, and the convergence degree of D-ABC and C-ABC algorithms was better than ABC algorithm. This is because the C-ABC and D-ABC algorithms introduced security-specific domain search strategies and information diffusion probabilistic model optimisation mechanisms that are more suitable for dealing with complex non-linear problems in solving the dual WoM diffusion product pricing problem. In particular, the CD-ABC algorithm, by combining the advantages of C-ABC and D-ABC, was able to adjust the search strategy more quickly, avoid premature convergence, and maintain a balance between exploration and exploitation. This allowed the algorithm to find solutions closer to the global optimum even when faced with the interaction of multiple factors in a changing market environment. For the same initial population, the test results of observing the influence of bees on the honey source search probability showed that some individuals in D-ABC had a high evolutionary frequency. When D was set to 30, the CD-ABC algorithm performed better than GABC, MABC, qABC, and dABC in 14 functions. In most functions, the CD-ABC showed efficient solution efficiency and rate of convergence. This is due to the efficient evolutionary strategy of the CD-ABC algorithm in dealing with the colony's search behaviour for nectar sources. In the D-ABC algorithm, individuals exhibited different evolutionary frequencies, and this differentiated evolutionary strategy provided more diverse search paths for the probabilistic model of information diffusion. When the evolution frequency parameter was set to 30, the CD-ABC algorithm was able to adjust its search strategy more accurately, which ensured excellent performance on various test functions, especially in comparison with other state-of-the-art algorithms such as GABC and MABC. This diversified search path not only accelerated the convergence speed but also improved the solution efficiency, enabling the CD-ABC algorithm to efficiently approximate the global optimal solution in complex optimisation problems, especially when simulating market decisions in real-world dynamic environments. The actual strategy solution findings showed that in the CBEC clothing pricing problem, the CD-ABC algorithm converged after 200 iterations, with lower iterations than common decision solving algorithms and much higher benefits than common algorithms, reaching 7.95e+11. The comprehensive imitation effect was -0.14, the WoM scale effect was 1.34, the pre-sale price was ultimately set at 347.49, and the sales price was 641.393. When solving the pricing problem of CBEC electronic products, after 230 iterations to reach convergence, the optimal profit decision obtained reached 5.31e+11, which was much higher than the results obtained by other algorithms. The research proposed a dual WoM diffusion product pricing model and the CD-ABC algorithm, which had better computational performance compared to common models in CBEC pricing decision-making problems. However, their lack of consideration for the background of the interaction of multiple factors in practical problems is also an area for further research to improve.

References

[1] Audrey, Guo. Cross-border E-commerce as the Main Force to Stabilize Foreign Trade. China's Foreign Trade, 2020, 581(05):53-55.

[2] Long X M, Chen Y J, Zhou J. Development of AR Experiment on Electric-Thermal Effect by Open Framework with Simulation-Based Asset and User-Defined Input. Artificial Intelligence and Applications. 2023, 1(1): 52-57.

[3] Arodh Lal Karn, Rakshha Kumari Karna, Bhavana Raj Kondamudi, Girish Bagale, Denis A. Pustokhin, Irina V. Pustokhina, Sudhakar Sengan. Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis. Electronic commerce research,2023,23(1):279-314.

[4] Apasrawirote D, Yawised,K, Muneesawang P. Digital marketing capability: the mystery of business capabilities. Marketing intelligence & planning, 2022, 40(4):477-496.

[5] Yan Z, Lu X, Chen Y, et al. Institutional distance, internationalization speed and cross-border e-commerce platform utilization. Management Decision, 2023, 61(1): 176-200.

[6] Li, Bing, Li, Jiahua, Ou, Xijun. Hybrid recommendation algorithm of cross-border e-commerce items based on artificial intelligence and Multiview collaborative fusion. Neural computing & applications, 2022, 34(9):6753-6762.

[7] Setkute J, Dibb S. 'Old boys' club': Barriers to digital marketing in small B2B firms. Industrial marketing management, 2022, 102(Apr.):266-279.

[8] Yan F, Taien L. The Effect of Performance Management System on Employee Productivity in Cross-Border E-Commerce Enterprises in China. Management research, 2022, 10(3):155-166.

[9] Li XL, Zhu, BH. Influence of environmental regulation on cross-border ecommerce export of agricultural products. Journal of environmental protection and ecology,2021,22(3):1347-1357.

[10] Ezaldeen H, Bisoy S K, Misra R, Alatrash R. Semantics-Aware Context-Based Learner Modelling Using Normalized PSO for Personalized E-learning. Journal of web engineering, 2022, 21(4):1187-1223.

[11] Cao L, Zhang W, Kan X, et al. A Novel Adaptive Mutation PSO Optimized SVM Algorithm for sEMG-Based Gesture Recognition. Scientific programming, 2021, 2021(Pt.5):823-836.

[12] Stojanovic B, Gajevic S, Kostic N, et al. Optimization of parameters that affect wear of A356/Al2O3 nanocomposites using RSM, ANN, GA and PSO methods. Industrial Lubrication and Tribology, 2022, 74(3):350-359.

[13] Jarndal A, Husain S, Hashmi M, et al. Large-Signal Modeling of GaN HEMTs Using Hybrid GA-ANN, PSO-SVR, and GPR-Based Approaches. IEEE Journal of the Electron Devices Society, 2021, 9195-208.

[14] Song JH, Kang SW, Kim YJ. Optimal design of the disc vents for high-speed railway vehicles using thermal-structural coupled analysis with genetic algorithm. Proceedings of the Institution of Mechanical Engineers, Part C. Journal of mechanical engineering science, 2022, 236(10):5154-5164.

[15] Barma M, Modibbo U M. Multiobjective mathematical optimization model for municipal solid waste management with economic analysis of reuse/recycling recovered waste materials. Journal of Computational and Cognitive Engineering, 2022, 1(3): 122-137.

[16] Drewitz, Alexander, Prevost, Alexis, Rodriguez, Pierre-Francois. Critical exponents for a percolation model on transient graphs. Inventiones Mathematicae, 2023, 232(1):229-299.

[17] Voskoglou M G. A Combined Use of Soft Sets and Grey Numbers in Decision Making. Journal of Computational and Cognitive Engineering, 2023, 2(1): 1-4.

[18] Maihulla A S, Yusuf I, Bala S I. Reliability and performance analysis of a series-parallel system using Gumbel–Hougaard family copula. Journal of Computational and Cognitive Engineering, 2022, 1(2): 74-82.

[19] Mohammad Hossein Heydari, Mahmood Haji Shaabani, Zahra Rasti. Orthonormal discrete Legendre polynomials for nonlinear reaction-diffusion equations with ABC fractional derivative and non-local boundary conditions. Mathematical Methods in the Applied Sciences, 2023, 46(12):13423-13435.

[20] Manpreet Kaur, Jagtar Singh Sivia. Artificial bee colony algorithm based modified circular-shaped compact hybrid fractal antenna for industrial, scientific, and medical band applications. International journal of RF and microwave computer-aided engineering, 2022, 32(3):32-44.

[21] Kayikci, Safak, Unnisa, Nazeer, Das, Anupam, Kanna, S. K. Rajesh, Murthy, Mantripragada Yaswanth Bhanu, Preetha, N. S. Ninu, Brammya, G. Deep Learning with Game Theory Assisted Vertical Handover Optimization in a Heterogeneous Network. International Journal of Artificial Intelligence Tools: Architectures, Languages, Algorithms, 2023, 32(4):12-45.

Planning and Expansion of the Transmission Network in the Presence of Wind Power Plants

Hui Sun*
Department of Automotive Engineering, Hebei Vocational University of Industry and Technology, Hebei 050091, China

***Abstract*—The proliferation of renewable energy sources, particularly wind farms, is rapidly gaining momentum owing to their numerous benefits. Consequently, it is imperative to account for the impact of wind farms on transmission expansion planning (TEP), which is a crucial aspect of power system planning. This article presents a multi-objective optimization model that utilizes DC load flow to address the TEP challenge while also incorporating wind farm uncertainties into the model. The present study aims to optimize the expansion and planning of the TEP in the power system by considering investment and maintenance costs as objective functions. To achieve this, a multi-objective approach utilizing the shuffled frog leaping algorithm (SFLA) is proposed and implemented. The proposed objectives are simulated on the RTS-IEEE 24-bus test network. The results obtained from the proposed algorithm are compared with those of the Genetic Algorithm (GA) to assess and validate the proposed approach.**

Keywords*—Wind farms; Transmission Expansion Planning (TEP); multi-objective optimization model; Shuffled Frog Leaping Algorithm (SFLA)

NOMENCLATURE

Abbreviations	
TEP	Transmission expansion planning
SFLA	Shuffled frog leaping algorithm
Parameters and variables	
Cij	The cost of new lines among buses i and j
Nij	The number of new lines among buses i and j
N0ij	The initial lines between buses i and j
lij	The length of new lines between buses i and j
CM, CO	The maintenance and operation costs
S	Intersection matrix for buses
f	Power flow vector
r	Resistance of lines
gij	Vector of the power generation
d	Load demand vector
fij	Power flow between lines in buses i and j
θi	Voltage angle in bus i

I. INTRODUCTION

A. Aims and Related Research

During the past few decades, problems have existed in the electric power generation sector in a traditional system, such as high production costs, environmental effects, high losses and low reliability [1]**.** Hence, distributed generation systems have been used in such systems, which are installed in the vicinity of consumption centers and have lower power, loss, and cost, as well as more reliability than traditional forms of electric energy generation. Wind energy is one of the energy sources that have received a lot of attention [2]. Investigating the effects of wind energy resources on power networks from various aspects, such as uncertainty and uncontrollability of generation power compared to conventional sources of energy production, has been studied as one of the important challenges in this field [3]. In addition to these cases, the distance from demand centers and the strong dependence of wind turbine production capacity on wind speed should be added as the biggest obstacles to the use of this energy [4]. The objectives of transmission expansion planning (TEP) in the power system are twofold: to plan power systems that ensure a dependable energy supply to customers and to identify optimal locations and methods for investing in new transmission lines that will facilitate reliable energy supply to customers. Additionally, TEP involves the operation of load growth based on demand forecasting [5]. Hence, to implement of reliable energy supply the increase the generation share of wind farms compared to the total generation of electric energy in the traditional systems [6], [7], it is necessary to have a reliable supply of energy in order to maintain customer satisfaction in the power system planning with regard to uncertainties [8], [9].

The research of the power systems considering different areas like power plants, transmission and distribution grids are assessed in this subsection. In research [10], energy optimization with TEP considering power generation of renewable energies such as solar power and wind turbines is proposed. Authors in study [11] installed and sized the solar panels and storage systems in the transmission lines with consideration of the power loss reduction reported. The operation of the generation units in the power plants by using the unit commitment approach is proposed in [12]. In (Moreira et al., 2017), the economic dispatch approach for energy scheduling in the power plants is used. The planning and operation modeling in the power grids with maximizing reliability is studied in [13]. The article in [14] presents an economic approach to power flow analysis, taking into account factors such as fuel costs in power plants and the operation of units during peak demand. Meanwhile, [15] models power flow with a focus on the cost of transmission lines and incorporates the use of Flexible AC Transmission Systems (FACTS) to enhance voltage index. The modeling economic of the microgrids for TEP and power grids for covering the uncertainty of renewable energy is studied in [16].

B. Contributions

This paper presents a multi-objective optimization model for TEP in the power system, taking into account the operation of wind farms. The objective functions based on investment and maintenance costs are modeled for implementing TEP in the power systems. The shuffled frog leaping algorithm (SFLA) is proposed for solving the optimization approach. The expansion of the wind farms in the TEP is modeled based uncertainty approach. The DC power flow is considered for TEP with wind farm operations. Hence, the contributions of this paper can be summarized as follows:

1) Proposing multi-objective modeling for TEP in the power system considering wind farm participation.

2) Implementing TEP by investment and maintenance costs.

3) Utilizing shuffled frog leaping algorithm (SFLA) for solving problems.

4) Modeling wind farm based on uncertainty approach.

II. TEP FORMULATION

The inability to accurately predict the load due to the uncertainty of power generation, such as wind energy, leads to the introduction of new technologies in electrical energy generation. Hence, wind farms, due to the randomness of the generation power in their generation, cannot be ignored. Therefore, the modeling of these uncertainties in the planning of power systems will lead to the creation of stronger planning that can meet different conditions [17]. It should be noted that the uncertainties of the network structure make the decision-making process difficult. In traditional planning, the main goal is to minimize investment costs [18]. However, in modern planning, several different goals are optimized simultaneously, and traditional methods are not able to provide acceptable solutions [19].

A. Wind power modeling

The amount of power produced by wind farms is contingent upon the speed of the wind. As a result, the power output of a wind turbine differs significantly from that of a conventional energy generation unit [20]. Hence, modeling wind power is formulated by the "power-speed" curve in Fig. 1 and other parameters of the turbine. Also, modeling wind turbines is formulated by Eq. (1) [20].

$$P_{WT}(v)=\begin{cases} 0 & if \quad v \le V_{cin} \\ P_{N,WT}\times\left(\dfrac{v-V_{cin}}{V_r-V_{cin}}\right) & if \quad V_{cin}\le v\le V_r \\ P_{N,WT} & if \quad V_r\le v\le V_{co} \\ 0 & if \quad V_{co}\le v \end{cases} \tag{1}$$

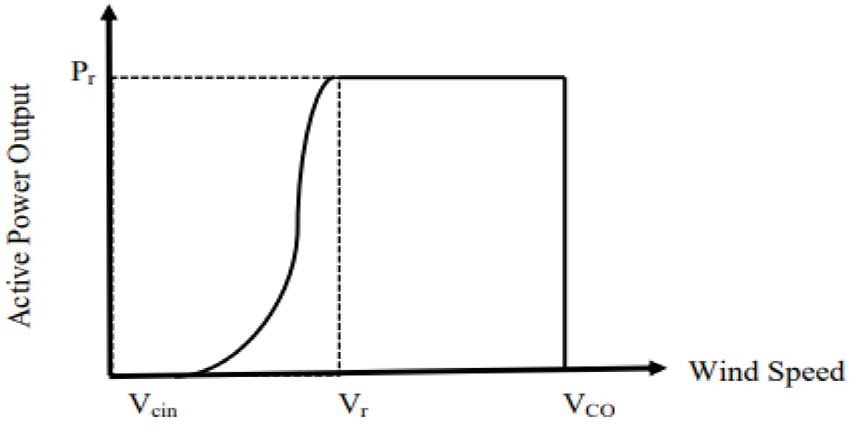


Fig. 1. Wind turbine "power-speed" characteristic.

B. Objective functions formulation

The objective functions, such as investment and maintenance costs, are minimized to TEP in this section. The modeling of the objectives is as follows:

1) Cost of investment: The modeling investment considering constraints for this objective is as follows:

$$\min f_{IC}=\sum_{i,j=1}^{n} C_{ij}\times l_{ij}\times N_{ij} \tag{2}$$

Subject to:

$$Sf+g=d \tag{3}$$

$$f_{ij}-r(N_{ij}^{0}+N_{ij})(\theta_i-\theta_j)=0 \qquad \forall i,j,n \tag{4}$$

$$\left|f_{ij}\right|\le(N_{ij}^{0}+N_{ij})f_{ij}^{\max} \qquad \forall i,j,n \tag{5}$$

$$0\le N_{ij}\le N_{ij}^{\max} \qquad \forall i,j,n \tag{6}$$

$$0\le g_{ij}\le g_{ij}^{\max} \qquad \forall i,j,n \tag{7}$$

Where constraints in Eq. (3) and Eq. (4) are power flow balance and power flow in branches *i* and *j*, respectively, the constraints in Eq. (5) and Eq. (6) limit the power flow and limit of the number lines *i* and *j*, respectively. Constraint in Eq. (7) is active power generation by units in buses *i* and *j*.

2) Maintenance and operation costs: The second objective is to minimize the maintenance and operation costs of the TEP:

$$\min f_2 = \sum_{i,j=1}^{n} C_M N_{ij} + C_O N_{ij} \tag{8}$$

In objective function in Eq. (8) first and second terms are maintenance and operation costs, respectively.

III. Optimization Method

This study uses SFLA as an optimization method to solve the objective functions. SFLA is modeled based on the population of frog groups or memeplexes to find food. This method addresses the creation of frog populations as part of local and global strategies and objective functional changes based on the replacement of existing frogs. SFLA can be done by following these steps [21], [22]:

1) Population generation: Randomly generates a population "p", considering each frog's position and search space.

2) Creation of the memeplexes: The frogs must be evenly distributed in the memeplexes, taking into account their fitness function, where the population of frogs as m memeplexes with n frogs is p = [m × n].

3) Update frog location: Frogs in memeplexes are updated based on their best and worst locations from local search. Then the worst frog (Ωw) is updated with the best frog (Ωb) in each memeplex and the whole memeplex with the best frog (Ωgb):

$$\Omega_w^{new} = \Omega_w + rand \times (\Omega_b - \Omega_w) \tag{9}$$

$$\Omega_w^{new} = \Omega_w + rand \times (\Omega_{gb} - \Omega_w) \tag{10}$$

Here, 0 ≤rand ≤1 is a random number. With (9), the worst frog Ω_w can be upgraded by placing the best frog Ω_b. In this step, if Ω_w^{new} is better than Ωw, Ωw is replaced by Ω_w^{new}; otherwise, Ωw can be replaced with Eq. (10) by the best frog in the Ω_{gb} memeplex. To find the optimal solution of SFLA, this process is performed for all iterations and memeplexes. The process of the SFLA is presented by Algorithm 1.

Algorithm 1: Process of the SFLA.

Algorithm 1 pseudocode of the SFLA

```
1. Start
2. Create population of P frogs, randomly;
    3. Calculate fitness function of the i frog;
    4. Sort the frogs based on their fitness;
    5. Distribution of the frogs by m memeplexe and n frog ($P = m×n$);
       In each memeplex;
      Determine $X_B$ and $X_W$;
          Improve the $X_W$ position using Eqs. (11) and (12);
      Repeat for number of iterations;
      Stopping critical satisfied?
      Yes
      End
      Else=go to step 2
```

taSince objectives are optimized in this study simultaneously. The frontier solutions will be obtained. The energy operator must determine the optimal solution for objectives in the frontier solutions as a decision maker. Hence, the max-min fuzzy method is proposed for a determined optimal solution as follows [23], [24]:

$$\Gamma\left(f_z(\vartheta)\right) = \begin{cases} 0 & otherwise \\ \dfrac{f_z^{\max} - f_z(\vartheta)}{f_z^{\max} - f_z^{\min}} & f_z^{\min} \le f_z(\vartheta) \le f_z^{\max} \\ 1 & f_z^{\min} \ge f_z(\vartheta) \end{cases} \tag{11}$$

$$\max\left\{\min \Gamma\left(f_z(\vartheta)\right)\right\} \tag{12}$$

Here, *Γ (f* —

Here, *Γ (f_z (ϑ))* and *$f_z(\vartheta)$* are membership functions or solutions in *zth* objective and value of objective at ϑth frontier solutions, respectively. Also, to determine the optimal solution in frontier solutions maximum and minimum procedure is presented in Eq. (11). In Eq. (12), a high rate of minimum solution is introduced as the optimal solution.

IV. Numerical Simulation

In this section, TEP studies have been carried out using the proposed algorithm in the MATLAB software environment in a system with a 4 GHz CPU and 6 GB RAM using DC load flow. The RTS-IEEE 24-bus test network is used for implementing TEP considering wind farm installation. In Fig. 2, the RTS-IEEE 24-bus test network with wind farms is shown. The wind speed based on average value is shown in Fig. 3. As should be mentioned, the TEP time study is considered for ten years. The average wind speed is considered in the simulation, and data from the wind farm is presented in Table I. The two wind farms have the same data. Also, load demand data is listed in Table II. Information on the generator units and test network are extracted from study [25]–[27].

TABLE I. Wind Farm Data

Parameters	Value
PN, WT	10 MW
Vr	10 m/s
Vcin	3 m/s
Vco	16 m/s
NWT	35

TABLE II. Load Demand Data

Bus	Demand (MW)	Bus	Demand (MW)
1	323	10	586
2	292	13	796
3	541	14	583
4	223	15	950
5	212	16	302

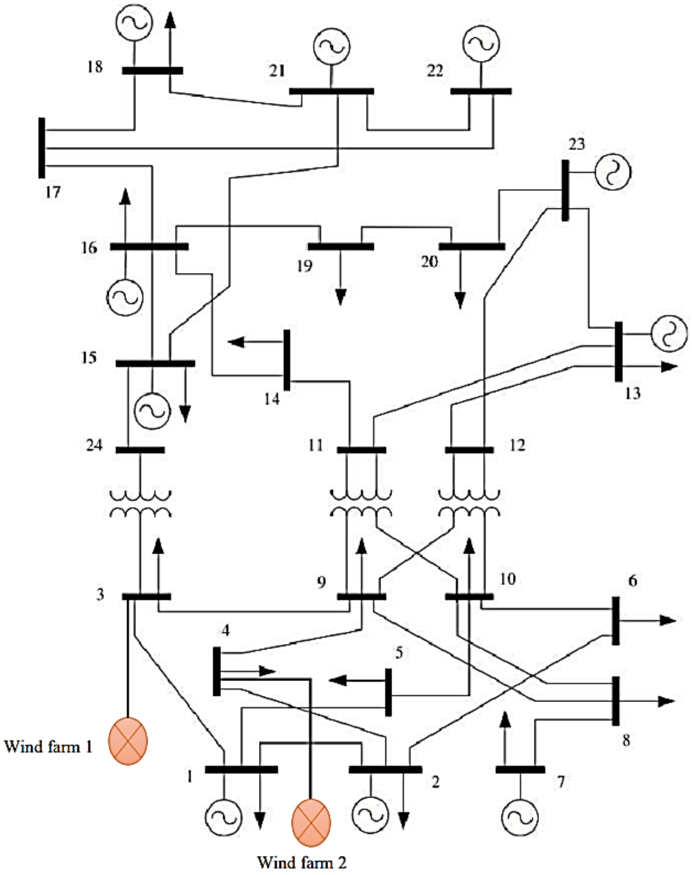


Fig. 2. RTS-IEEE 24-bus test network with wind farms.

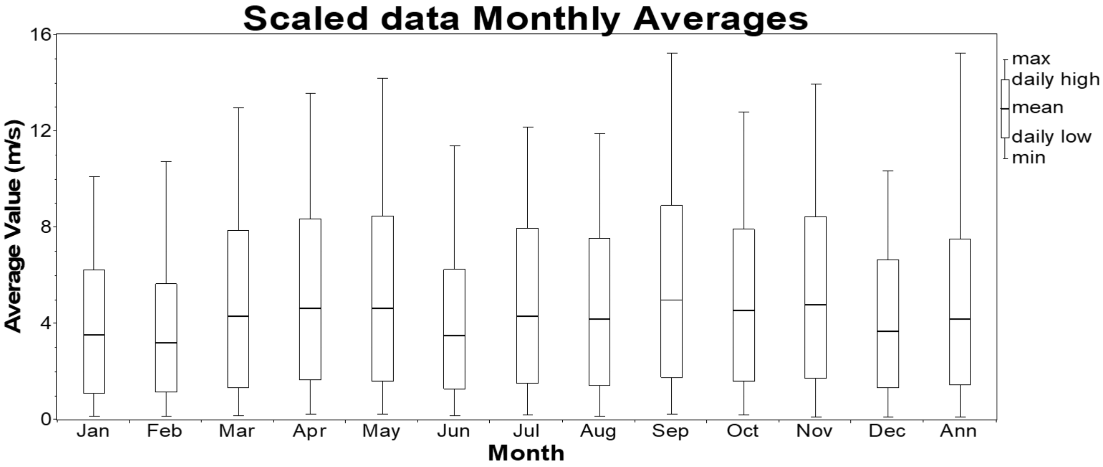


Fig. 3. Wind speed in test network.

A. *Results Analyse*

To examine the impact of various conditions on the outcomes of resolving the TEP issue utilizing the suggested algorithm, the ensuing scenarios were analyzed and executed on the 24-bus RTS-IEEE test system. The scenarios are as follows:

Scenario A) Implementing TEP without wind farms.

Scenario B) Implementing TEP with wind farms.

Also, in this study, the proposed optimization approach is compared with the Genetic Algorithm (GA) for verification and confirmation of the SFLA. In Fig. 4 and 5, frontier solutions of the objective functions for scenarios A and B by comparing SFLA with GA are shown, respectively. The obtained optimal solution by fuzzy method for the first and second objectives by SFLA in scenario A are equal to

$11996.3 and $768.6, respectively. The optimal solution generated by GA yields a first objective amount of $12453.3 and a second objective amount of $796.4. These results of the SFLA represented more convergence of the optimization for TEP in scenario A than GA. The value of the optimal solutions in scenario A, by the fuzzy method for SFLA and GA, is equal to 0.46 and 0.43, respectively.

On the other side, with implementing TEP with wind farms, the results of the objective functions in Fig. 5 are more optimizer than scenario. In scenario B, the SFLA algorithm has yielded optimal solutions of $11153.4 and $750.6 for the first and second objective functions, respectively. It's visible with the installation of the wind farms; expansion of the transmission lines for supply load demand is optimized than scenario A. Furthermore, the utilization of GA in scenario B results in a decrease in the values of both the first and second objective functions. These reductions of the objective functions in scenario B are due to more generation capacities, dropping power flow in lines, and increasing line capacities in meeting load demand.

Fig. 6 and Fig. 7 depict TEP implementation in scenarios A and B using SFLA and GA in the RTS-IEEE 24-bus test network. The orange lines represent optimal TEP solutions to meet load demand while considering economic power generation from the units. The implementation of the TEP by SFLA in both scenarios leads to reduce investment costs and maintenance and operation costs in comparison with GA.

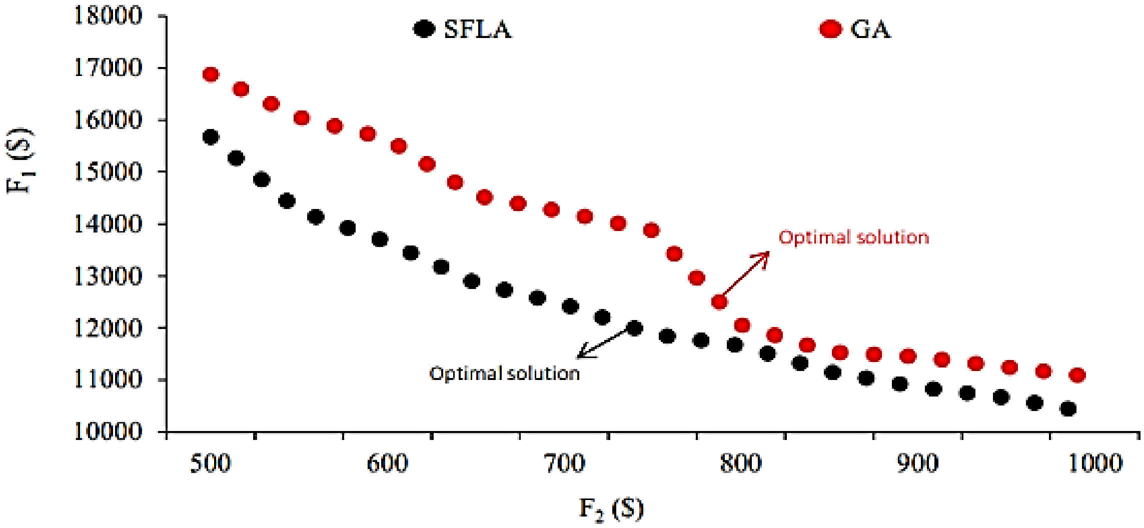


Fig. 4. Objectives in scenario A.

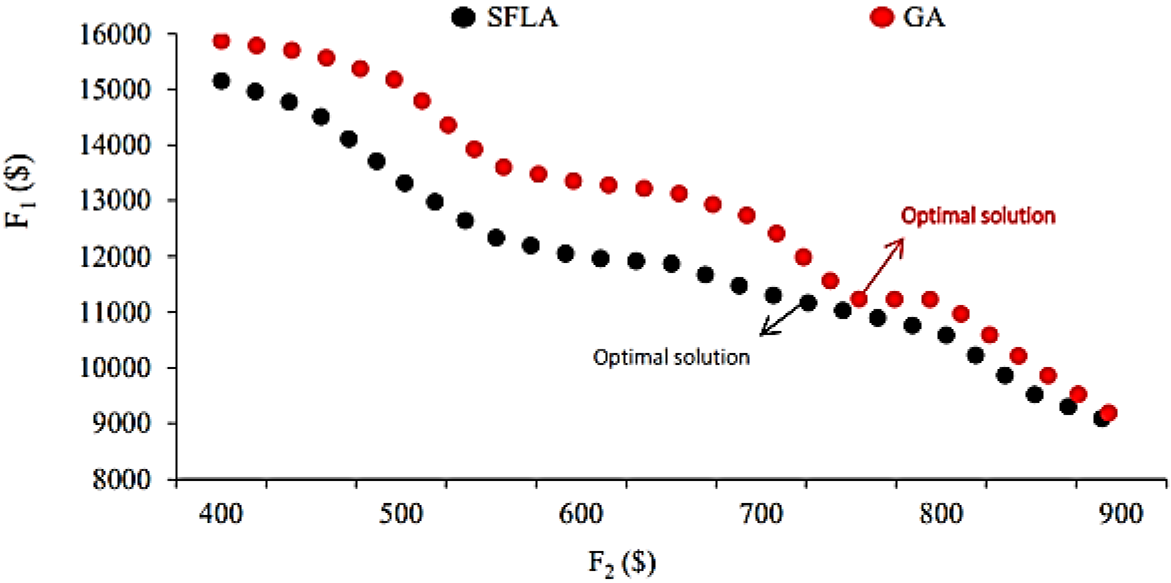
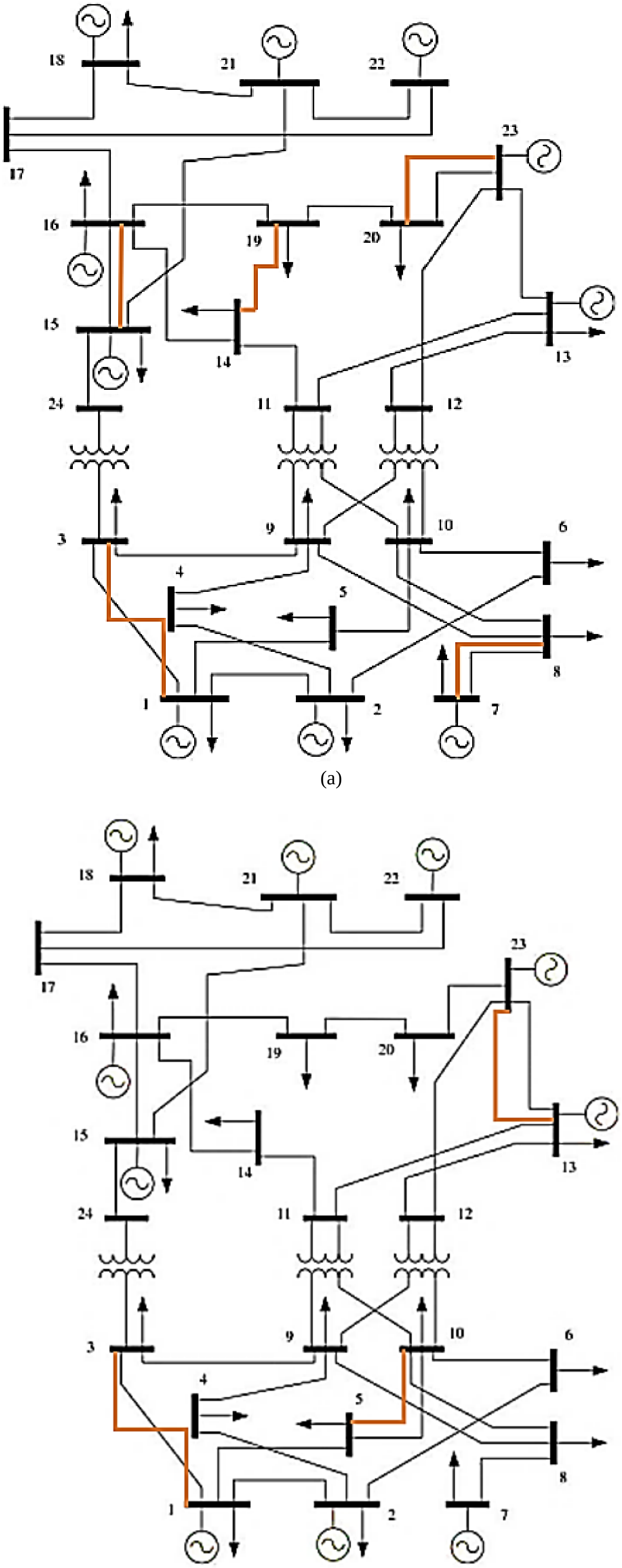


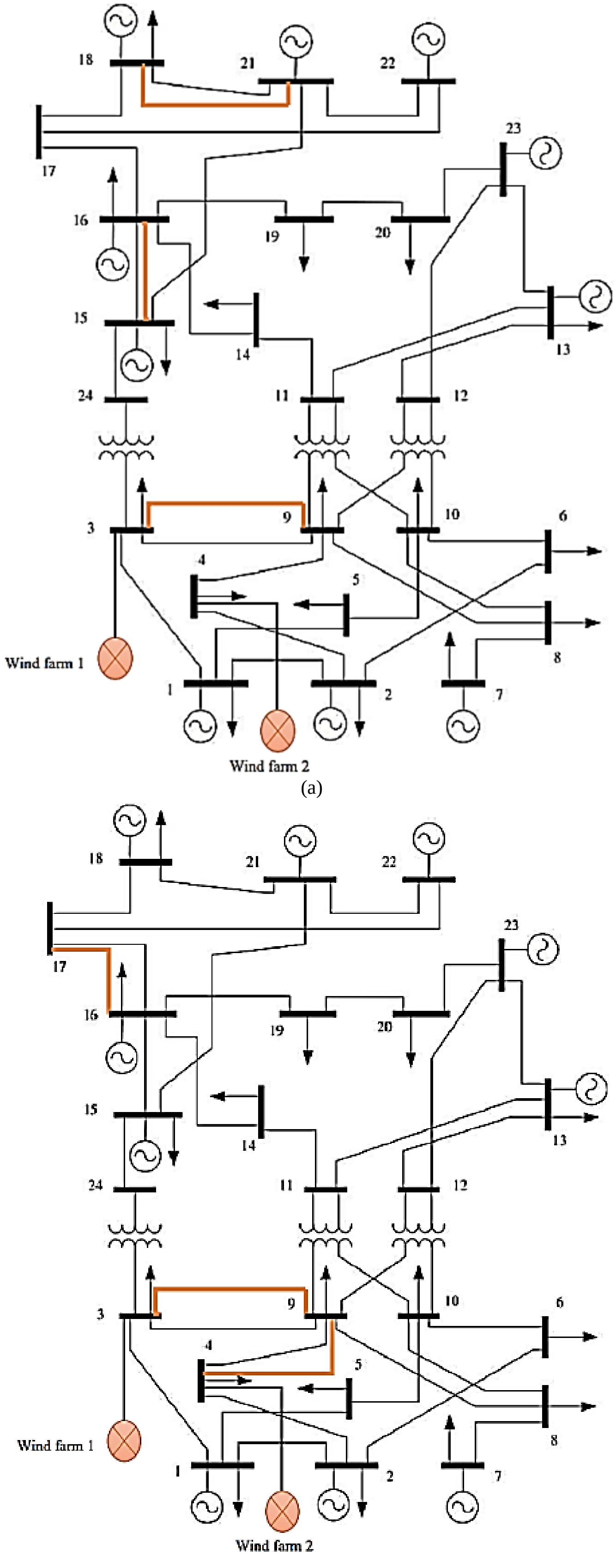
Fig. 5. Objectives in scenario B.



(a)

(b)

Fig. 6. TEP in scenario A. a) GA and b) SFLA.



(a)

(b)

Fig. 7. TEP in scenario B. a) GA and b) SFLA.

V. Conclusion

As a result of the escalating load growth and the integration of renewable resources into the power system, TEP has become an unavoidable issue. This article addresses the TEP problem as a multi-objective optimization problem, taking into account the presence of wind farms. Specifically, the study investigates the effectiveness of the SFLA on the modified RTS-IEEE 24-bus network. The proposed method aims to minimize investment and maintenance costs. Comparative analysis between SFLA and GA demonstrates the superiority of the former in achieving the desired objectives.

In comparison to other models, a significant advantage of this particular model lies in its investment and maintenance costs, which align with the fundamental objectives of TEP. Consequently, it is imperative to incorporate the cost of investment and maintenance as a component of fixed costs, given their crucial role in the planning process of power systems. Furthermore, the inclusion of wind farms in TEP-related matters will progressively enhance the performance of power networks in accordance with the load demand rate.

References

[1] Han, L., & Yu, H.-H. (2023). An empirical study from Chinese energy firms on the relationship between executive compensation and corporate performance. Nurture, 17(3), 378–393. https://doi.org/10.55951/nurture.v17i3.356.

[2] Rehan, R. . (2022). Investigating the capital structure determinants of energy firms. Edelweiss Applied Science and Technology, 6(1), 1–14. https://doi.org/10.55214/25768484.v6i1.301.

[3] Lak Kamari, M., H. Isvand, and M. Alhuyi Nazari. "Applications of multi-criteria decision-making (MCDM) methods in renewable energy development: A review." Renewable Energy Research and Applications 1.1 (2020): 47-54.

[4] Molamohamadi, Z., and M. R. Talaei. "Analysis of a proper strategy for solar energy deployment in Iran using SWOT matrix." Renewable Energy Research and Applications 3.1 (2022): 71-78.

[5] Beiranvand, A., et al. "Energy, exergy, and economic analyses and optimization of solar organic Rankine cycle with multi-objective particle swarm algorithm." Renewable Energy Research and Applications 2.1 (2021): 9-23.

[6] Salek, Farhad, et al. "Investigation of Solar-Driven Hydroxy gas production system performance integrated with photovoltaic panels with single-axis tracking system." Renewable Energy Research and Applications 3.1 (2022): 31-40.

[7] Norouzi, N., & Bozorgian, A. (2023). Energy and exergy analysis and optimization of a Pentageneration (cooling, heating, power, water and hydrogen). Iranian Journal of Chemistry and Chemical Engineering.

[8] Norouzi, N., Ebadi, A. G., Bozorgian, A., Hoseyni, S. J., & Vessally, E. (2021). Energy and exergy analysis of internal combustion engine performance of spark ignition for gasoline, methane, and hydrogen fuels. Iranian Journal of Chemistry and Chemical Engineering, 40(6), 1909-1930..

[9] Jovijari, F., Kosarineia, A., Mehrpooya, M., & Nabhani, N. (2023). Exergy, Exergoeconomic and Exergoenvironmental Analysis in Natural Gas Liquid Recovery Process. Iranian Journal of Chemistry and Chemical Engineering, 42(1), 237-268.

[10] N. G. Ude, H. Yskandar, and R. C. Graham, “A comprehensive state-of-the-art survey on the transmission network expansion planning optimization algorithms,” IEEE Access, vol. 7, pp. 123158–123181, 2019.

[11] Abed Almoussaw, Z., Abdul Karim, N., Taher Braiber, H., Ali Abd Alhasan, S., Raheem Alasadi, S., H. Ali Omran, A., Tariq Kalil, Z., V. Pavlova, I., & Alabdallah, Z. (2022). Analysis of geothermal energy as an alternative source for fossil fuel from the economic and environmental point of view: A case study in Iraq. Caspian Journal of Environmental Sciences, 20(5), 1127-1133. doi: 10.22124/cjes.2022.6093.

[12] Á. García-Cerezo, R. García-Bertrand, and L. Baringo, “Enhanced representative time periods for transmission expansion planning problems,” IEEE Transactions on Power Systems, vol. 36, no. 4, pp. 3802–3805, 2021.

[13] Shevchenko, V., Soloviev, A., & Popova, N. (2021). Energy and economic efficiency of corn silage production with flat grain of soy bean on reclaimed lands of upper volga. Caspian Journal of Environmental Sciences, 19(5), 947-950. doi: 10.22124/cjes.2021.5272.

[14] Y. Wang et al., “Transmission network dynamic planning based on a double deep-Q network with deep ResNet,” IEEE Access, vol. 9, pp. 76921–76937, 2021.

[15] A. Arabali, M. Ghofrani, M. Etezadi-Amoli, M. S. Fadali, and M. Moeini-Aghtaie, “A multi-objective transmission expansion planning framework in deregulated power systems with wind generation,” IEEE Transactions on Power Systems, vol. 29, no. 6, pp. 3003–3011, 2014.

[16] A. Khodaei and M. Shahidehpour, “Microgrid-based co-optimization of generation and transmission planning in power systems,” IEEE transactions on power systems, vol. 28, no. 2, pp. 1582–1590, 2012.

[17] G. C. Oliveira, S. Binato, and M. V. F. Pereira, “Value-based transmission expansion planning of hydrothermal systems under uncertainty,” IEEE Transactions on power systems, vol. 22, no. 4, pp. 1429–1435, 2007.

[18] J. A. Aguado, S. Martin, C. A. Pérez-Molina, and W. D. Rosehart, “Market Power Mitigation in Transmission Expansion Planning Problems,” IEEE Transactions on Energy Markets, Policy and Regulation, 2023.

[19] F. Chen, J. Liu, M. Zhao, and H. Liu, “Congestion Identification and Expansion Planning Methods of Transmission System Considering Wind Power and TCSC,” IEEE Access, vol. 10, pp. 89915–89923, 2022.

[20] Bakhshipour, A., Bagheri, I., Psomopoulos, C., & Zareiforoush, H. (2021). An overview to current status of waste generation, management and potentials for waste-to-energy (Case study: Rasht City, Iran). Caspian Journal of Environmental Sciences, 19(1), 159-171. doi: 10.22124/cjes.2021.4506.

[21] M. Eusuff, K. Lansey, and F. Pasha, “Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization,” Engineering optimization, vol. 38, no. 2, pp. 129–154, 2006.

[22] A. Alghazi, S. Z. Selim, and A. Elazouni, “Performance of shuffled frog-leaping algorithm in finance-based scheduling,” Journal of computing in civil engineering, vol. 26, no. 3, pp. 396–408, 2012.

[23] S. Dorahaki, A. Abdollahi, M. Rashidinejad, and M. Moghbeli, “The role of energy storage and demand response as energy democracy policies in the energy productivity of hybrid hub system considering social inconvenience cost,” J Energy Storage, vol. 33, p. 102022, 2021.

[24] H. Chamandoust, G. Derakhshan, S. M. Hakimi, and S. Bahramara, “Tri-objective scheduling of residential smart electrical distribution grids with optimal joint of responsive loads with renewable energy sources,” J Energy Storage, vol. 27, p. 101112, 2020.

[25] R. Minguez, R. García-Bertrand, J. M. Arroyo, and N. Alguacil, “On the solution of large-scale robust transmission network expansion planning under uncertain demand and generation capacity,” IEEE Transactions on Power Systems, vol. 33, no. 2, pp. 1242–1251, 2017.

[26] C. Ordoudis, P. Pinson, J. M. Morales, and M. Zugno, “An updated version of the IEEE RTS 24-bus system for electricity market and power system operation studies,” Technical University of Denmark, vol. 13, 2016.

[27] R. Ucheniya, A. Saraswat, S. A. Siddiqui, S. K. Goyal, and N. Kanwar, “A wind farm modeling in IEEE-24 bus reliability test system on DIgSILENT power factory,” in 2020 International Conference on Intelligent Engineering and Management (ICIEM), IEEE, 2020, pp. 477–483.

LPDA: Cross-Project Software Defect Prediction Approach via Locality Preserving and Distribution Alignment

Jin Xian[1], Jinglei Li[2]*, Quanyi Zou[3], Yunting Xian[4]
School of Computer Science and Engineering South China University of Technology Guangzhou 510006, China[1,4]
The Second Branch of China Railway Electrification Engineering Bureau Group Co., Ltd Guangzhou 511492, China[2]
School of Journalism and Communication South China University of Technology Guangzhou 510006, China[3]
Guangzhou 510006, China Guangdong Yousuan Technology Co., Ltd, FoShan 528000, P.R.China[4]

Abstract*—Cross-Project Defect Prediction (CPDP) based on domain adaptation aims to achieve defect prediction tasks in an unlabeled target software project by borrowing the defect knowledge extracted from well-annotated source software projects. Most existing CPDP approaches enhance transferability between projects but struggle with misalignments due to limited exploration of class-specific features and inability to preserve original local relationships in transformed features. In order to tackle these challenges, The article introduces a novel Cross-Project Defect Prediction (CPDP) approach called Local Preserving and Distribution Alignment (LPDA). This approach addresses the challenge of misalignments in CPDP due to limited exploration of discriminative feature representations and the failure to preserve original local relationship consistency. LPDA combines transferability and discriminability for CPDP tasks. It uses locality-preserving projection to maintain module consistency and distribution alignment, which includes transferable and discriminant distribution alignment. The former narrows the distributions of both source and target projects, while the latter increases the discrepancy between different classes across projects. The effectiveness of LPDA was tested through 118 cross-project prediction tasks involving 22 software projects from four distinct repositories. The results showed that LPDA outperforms baseline CPDP methods by efficiently learning representations that integrate transferability and discriminability while preserving local geometry to optimize distances within and between categories.

Keywords*—*Cross project defect prediction; discriminative distribution alignment; local preserving; domain adaption

I. Introduction

As software grows in size and complexity, defects inevitably arise, compromising quality and security [1], [2]. Ensuring software quality is therefore crucial before its release [3]. Software Defect Prediction (SDP) is a key technique to improve reliability by identifying potential defects in software modules, allowing for better allocation of testing resources. This technique uses historical data, like past source code and defect reports, to build models that can predict defects in new modules. Common methods for creating these models include neural networks [4], Naive Bayes [5], and support vector machines [6]. When prediction models are based on data from the same project, it's known as Within Project Defect Prediction (WPDP) [7], [8]. However, not all companies maintain historical defect data, and for those scenarios, Cross Project Defect Prediction (CPDP) uses data from external projects to build prediction models [9], [10].

CPDP aims to achieve defect prediction tasks in an unlabeled target project by learning the defect knowledge obtained from a source software project [11], [12]. However, Defect prediction for a target project is challenging due to differences from the source project, such as coding languages and developer expertise, which prevent direct knowledge transfer [13]. Domain adaptation is used to bridge the gap between projects, allowing defect knowledge to be transferred by adjusting features or instances in CPDP methods [14]. Various transfer learning algorithms from the literature [15], [16], [17], [18] are used to narrow marginal and/or conditional distribution difference between two projects. The CPDP approaches based on instance level select or reweight appropriate instances to decline the unfortunate impact from irrelevant cross-project data. For example, Ma et al. [19] proposed the Transfer Naive Bayes (TNB), introducing data gravitation that reweights instances of the source project. Moreover, software defect data often exhibits class-imbalance, with a significantly larger number of non-defective instances compared to defective instances. Class-imbalance has been broadly investigated both in WPDP [20], [21] and CPDP [22], [23], [24]. The impact of imbalanced datasets on the ranking of the approaches is also assessed [25], which addresses both class-imbalance and distribution mismatching. Tong et al. [11] introduced KSETE (Kernel Spectral Embedding Transfer Ensemble) to tackle class-imbalance in both homogeneous cross-project and heterogeneous cross-project scenarios.

Existing CPDP methods effectively reduce the gap between source and target projects, yet they overlook class distinction and disrupt local instance relationships. This can blur the decision boundary and misplace instances in the feature space, making accurate predictions difficult. Integrating locality-preserving techniques with domain adaptation could improve performance by maintaining the original data structure. Locality preserving projection is a typical approach based on manifold learning [26] that allows for learning a favourable feature space where the local consistency in the raw feature space can be effectively retained. The perfect performance will be obtained by integrating locality preserving projection and domain adaptation [27], [28], [29], but Locality-preserving objectives have not received thorough exploration within the CPDP domain.

To overcome the aforementioned limitations, this article introduces a new CPDP method called Local Preserving and Distribution Alignment (LPDA), which focuses on maintaining class distinctions and local data relationships. It aligns distributions globally and locally while keeping instances of the same class close and separating different classes. Additionally, new representations are generated in a low-dimensional subspace where instances instances from the same class remain closely, while instances from different classes are positioned farther apart. Extensive CPDP tasks on 22 open-source projects from four software repositories validate the effectiveness of the proposed approaches. The evaluation metrics used included F-measure, Balance, MCC, and AUC. The Wilcoxon signed rank test and Scott-Knott ESD test were adopted to statistical significance test. The contributions of this article can be summarised as follows:

1) Our proposed CPDP method, unlike previous ones, enhances transferability by reducing distribution discrepancies and also focuses on discriminability between different classes in the projects..
2) In CPDP, using locality-preserving projection maintains local consistency within classes, keeping similar instances closer and distinctly separating different categories.
3) The extensive experiments on 22 software projects from four software repositories demonstrate that the proposed LPDA approach is superior to several state-of-the-art CPDP approaches in terms of four performance indicators.

The article is organized as follows. Section II reviews prior works on CPDP and subspace alignment-based domain adaptation. Section III introduces the technical specification of our LPDA approach. Section IV covers the experiment configuration, encompassing aspects like open datasets, statistical tests, evaluation criteria, and research questions. In Section V, Extensive experiments along with analysis under CPDP scenarios are presented in detail.

II. Related Work

In this section, we will provide a concise overview of previous research in the domains of cross-project defect prediction and domain adaptation based on feature alignment.

A. Cross Project Defect Prediction

CPDP aims to seek instances with a high likelihood of defects in an unlabeled target software project using a predictive model trained on other well-labeled software projects[30]. In earlier research, Zimmermann et al. [31] investigated different factors that might influence the cross-project prediction performance for the first time. They have found that a few tasks (only 3.4% of the cross-project tasks) could achieve adequate prediction results and CPDP tasks between the projects are not symmetrical. The current CPDP approaches can be broadly categorized into two groups: homogeneous cross-project and heterogeneous cross-project, based on the similarity of software metrics (features) [11], [32]. In the context of homogeneous cross-project, the source and target projects share the same feature space, whereas in heterogeneous cross-project, the features of source and target projects differ. The CPDP approach employed in this paper falls under the category of homogeneous cross-project.

According to the theory of knowledge transfer, the current CPDP approaches can be mainly divided into instance transferring and feature transferring. The CPDP approaches based on instance transferring seek to select or reweight relevant instances from the source project data, which can be advantageous for the target task. The CPDP approaches based on feature transferring pay attention to learning shared feature representations for the two projects so as to narrow the distribution discrepancy between them. The former either chooses or adjusts the training data to mitigate the influence of detrimental information. The latter ensures that the source and the target projects exhibit a comparable distribution within the newly created feature subspace.

To the best of our knowledge, Turhan et al. [33] presented the CPDP approach that focus on instance transferring named Nearest Neighbor filter (NN-filter). They used KNN to gather close instances together to construct a similar training dataset.

In greater detail, for every instance of target project, NN-filter method picks the ten nearest instances from the source project and subsequently incorporates them into the training dataset. Building upon the NN-filter, Peters et al. [34] presented the Peter-filter approach, which selected instances using the k-means clustering algorithm. Using different clustering algorithms to select instances, Kanwata et al. [35] and Bhat et al. [36] presented two different CPDP approaches. Lately, Hosseini et al. [37] presented a search-based genetic instance selection approach, using genetic algorithm to select training data. These CPDP approaches based on instance selecting led to the source project wasting some data or a few available instances. To solve these problems, Ma et al. [19] proposed TNB, using the concept of data gravity, the transfer weights from the source projects is computed. These weights could strengthen the instances with significant correlations and weaken the impact of ineffective instances.

Drawn from transfers component analysis (TCA) [38], a classical transferring learning algorithm, Nam et al. [17] proposed the TCA+. This method adds normalization rules to process source and target data before distribution alignment. Liu et al. [39] detected that the prediction performance of TCA+ is erratic. Thus, they proposed two-phase transfer learning (TPTL) approach to m the issue of instability. TCA+ and TPTL only narrow the disparity in marginal distribution between the source and target projects. Simultaneously considering both the marginal and conditional distributions, Qiu et al.[15] and Xu et al. [18] proposed joint distribution matching (JDM) and balanced distribution adaptation (BDA), respectively.

Since deep learning has become capable of automatically extracting semantic features from ASTs of software program, many researchers have used it in the research SDP. Wang et al. [40] claimed that only used traditional software metrics were far from enough and represented the relationship between semantic features and software programs by abstract syntax tree (AST). Then they applied a deep belief network (DBN) to obtain software code semantic features from ASTs. Subsequently, Li et al. [41] constructed a extracting feature approach based on CNN to extract semantic features, and then integrated this features and traditional software metrics to address the absence of semantic information. However, these shallow learning and deep learning CPDP approaches consider only the transferability between source and target projects but ignore discriminability between classes.

B. Feature Alignment-based Domain Adaptation

Feature alignment-based domain adaptation approaches aim to learn a low-dimensional feature subspace where the disparity in distribution between the source and target data is explicitly narrowed. As a pioneer, Pan et al. [38] proposed to map both source and target data into a shared feature subspace via TCA, which aligns the marginal distributions between the source and target domains while maximizing the data variance in adaptation process. However, focusing only on aligning marginal distributions is insufficient for better learning purposes. Therefore, Long et al. [42] proposed to align both marginal and conditional distributions between two domains via joint distribution analysis (JDA). Similarly, Wang et al. [43] proposed a balanced domain adaptation (BDA), which simultaneously aligns the marginal and conditional distribution discrepancy and exploits a balance factor to adjust their importance degrees. JDA and BDA employ the classifier trained on source domain data to generate pseudo labels for the unlabeled target domain. After that, many variants will emerge. For example, Wang et al. [44] further proposed manifold embedded distribution alignment (MEDA) to dynamically adjust the relative importance of these two distributions. Zhao et al. [45] proposed discriminative joint probability MMD (DJP-MMD), which not only minimizes the divergence in joint distribution between two domains, but also maximizes the divergence in joint probability distribution between distinct classes in different domains to learn discriminative feature representation.

III. Research Methodology

In this section, we will present the LPDA approach for CPDP in more details. After describing the descriptions of notations used in this article.

A. Notations

$\mathcal{D}_s = \{\mathbf{x}_i^s, y_i^s\}_{i=1}^{n_s}$ and $\mathcal{D}_t = \left\{\mathbf{x}_j^t\right\}_{j=1}^{n_t}$ respectively represent the source project and the target project under the assumptions that both marginal distributions and conditional distributions of source and target projects are inequality($P\left(\mathbf{x}^s\right) \neq P\left(\mathbf{x}^t\right)$ and $Q\left(y^s \mid \mathbf{x}^s\right) \neq Q\left(y^t \mid \mathbf{x}^t\right)$). Let $\mathbf{X}_s \in \mathbb{R}^{d \times n_s}$ ($\mathbf{X}_t \in \mathbb{R}^{d \times n_t}$) is the source project data matrix (target project data matrix) containing $n_s(n_t)$ instance with d-dimension. $\mathbf{X}^c$ denotes a set of instances with label c. $\mathbf{x}_i^s \in \mathbf{X}_s$ and $\mathbf{x}_j^t \in \mathbf{X}_t$ are the $i-th$, $j-th$ instances in the source project and target project, respectively. The proposed CPDP approach aims to learn a transformation matrix $\mathbf{A}$ to transform the instances from the original feature space into a low-dimensional subspace where the distributions can be aligned and the important properties of the original data can be preserved. M_{mar} and M_{con} denote the distribution matching of marginal and the conditional distributions, respectively.

B. Overall Framework of LPDA

The previous CPDP approaches, such as JDM [15], TCA+ [17], TPTL [39] and BDA [18], learn global feature representation to narrow the distribution gap between different projects for the purpose of higher transferability, but disregard the discriminability between different classes. Additionally, these CPDP approaches cannot well preserve original local relationship consistency instances of shared the same label after transforming the features. To address these problems, this article proposes a novel CPDP approaches called Locality Preserving and Distribution Alignment (LPDA). Peculiarly, LPDA tries to fulfill three complementary objectives as follows. (1)

1) It aims at the characteristics of cross-project significant variations and draws from the idea of domain adaptation to reduce the global inter-project difference and the local intra-class discrepancy for to increaser transferability.
2) For discriminability between defective and non-defective classes, LPDA leverages class-wise MMD to enlarge the distance of different classes.
3) LPDA introduces the reward graph and penalty graph to maximize preserve the geometric structure of project instances.

The overall objective function of LPDA is follows:

$$\begin{aligned}\underset{\mathbf{A}}{\mathcal{L}} = \arg\min_{\mathbf{A}} & \underbrace{\mathcal{M}_t\left(\mathbf{X}_s, \mathbf{X}_t, \mathbf{A}\right)}_{\text{transferability}} - \mu \underbrace{\mathcal{M}_d\left(\mathbf{X}_s, \mathbf{X}_t, \mathbf{A}\right)}_{\text{discriminability}} \\ & + \eta \underbrace{\mathcal{G}\left(\mathbf{X}_s, \mathbf{X}_t, \mathbf{A}\right)}_{\text{geometric structure}} + \lambda \underbrace{\Omega(\mathbf{A})}_{\text{regularization}} .\end{aligned} \tag{1}$$

In Eq. (1), $\mathcal{M}_t$ is distribution difference in both marginal and conditional distributions across two project. $\mathcal{M}_d$ implies the distribution difference of different class between different projects. $\mathcal{G}$ represents the term of manifold regularization. Ω is the term structural risk. In addition μ, η and λ are the trade-off parameters.

C. Transferability in LPDA

Drawing on previous literatures [15], [18], we adopt maximum mean discrepancy (MMD) [38] to measure both the marginal distributions and conditional distribution distances between source and target projects. The marginal distributions distance can be achieved as follows:

$$\begin{aligned} M_{\text{mar}}\left(\mathbf{X}_s, \mathbf{X}_t\right) = & \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{x}_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(\mathbf{x}_j^t) \right\|_{\mathcal{H}}^2 \\ & = \operatorname{tr}(\mathbf{A}^\top \mathbf{K} \mathbf{M}_0 \mathbf{K}^\top \mathbf{A}), \end{aligned} \tag{2}$$

where, $\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_t]$. $\mathbf{M}_0$ is the marginal distribution MMD matrix and it is computed as

$$(\mathbf{M}_0)_{ij} = \begin{cases} \frac{1}{n_s^2}, & \text{if} \quad \mathbf{x}_j, \mathbf{x}_i \in \mathbf{X}_s, \\ \frac{1}{n_s^2}, & \text{if} \quad \mathbf{x}_j, \mathbf{x}_i \in \mathbf{X}_t, \\ \frac{-1}{n_s n_t}, & \quad \text{otherwise.} \end{cases} \tag{3}$$

Pseudo-labels in the target project are annotated by a classifier trained on the source project to represent conditional distribution. The conditional distributions distance can be achieved as follows:

$$M_{\mathrm{con}}\left(\mathbf{X}_s, \mathbf{X}_t\right) = \sum_{c=1}^{C} \left\| \frac{1}{n_{s,c}} \sum_{i=1}^{n_{s,c}} \phi(\mathbf{x}_i^{s,c}) - \frac{1}{n_{t,c}} \sum_{j=1}^{n_{t,c}} \phi(\mathbf{x}_j^{t,c}) \right\|_{\mathcal{H}}^2 = \mathrm{tr}(\mathbf{A}^\top \mathbf{K} \sum_{c=1}^{C} \mathbf{M}_c \mathbf{K}^\top \mathbf{A}), \tag{4}$$

where, $n_{s,c}$ and $n_{s,t}$ refer to the numbers of the c class instances of source and target projects, respectively. $\mathbf{M}_c$ is the conditional distribution MMD matrix with the label c, and it is computed as

$$(\mathbf{M}_c)_{ij} = \begin{cases} \frac{1}{n_{s,c}^2}, & \text{if} \quad \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_s^c, \\ \frac{1}{n_{t,c}^2}, & \text{if} \quad \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_t^c, \\ \frac{-1}{n_{s,c} n_{t,c}}, & \text{if} \left\{ \begin{array}{l} \mathbf{x}_i \in \mathbf{X}_s^c, \mathbf{x}_j \in \mathbf{X}_t^c, \\ \mathbf{x}_i \in \mathbf{X}_t^c, \mathbf{x}_j \in \mathbf{X}_s^c, \end{array} \right. \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Here, M_{mar} plus M_{con} is rewritten as the follows:

$$\mathcal{M}_t\left(\mathbf{X}_s, \mathbf{X}_t, \mathbf{A}\right) = M_{\mathrm{mar}} + M_{\mathrm{con}} = \mathrm{tr}(\mathbf{A}^\top \mathbf{X} \sum_{c=0}^{C} \mathbf{M}_c \mathbf{X}^\top \mathbf{A}). \tag{6}$$

This section considers the transferability between different projects from both marginal and conditional distributions perspectives. However, only considering the transferability might not be sufficient for better prediction performance.

D. Discriminability in LPDA

In this section, we explore the discriminability between defective and no-defective classes by leveraging class-wise MMD to augment the distribution distance between different class, which can be achieved as follows:

$$\mathcal{M}_d\left(\mathbf{X}_s, \mathbf{X}_t, \mathbf{A}\right) = \sum_{c=1}^{C} \sum_{\widehat{c} \neq c} \left\| \frac{1}{n_{s,c}} \sum_{i=1}^{n_{s,c}} \phi(\mathbf{x}_i^{s,c}) - \frac{1}{n_{t,\widehat{c}}} \sum_{j=1}^{n_{t,\widehat{c}}} \phi(\mathbf{x}_j^{t,\widehat{c}}) \right\|_{\mathcal{H}}^2. \tag{7}$$

In order to facilitate the calculation, we introduce a one-hot coding label matrix to calculate the class-wise MMD based on the literature [45]. In particular, the source project and target project one-hot coding label matrices are written as $\mathbf{Y}_s = [c_{s,1}, ..., c_{s,n_s}]$ and $\widehat{\mathbf{Y}}_t = [c_{s,1}, ..., c_{s,n_s}]$. Two matrices are defined as follows

$$\mathbf{U}_s = \frac{1}{n_s} \left[\mathbf{Y}_s(:,1) \bullet (C-1), \ldots, \mathbf{Y}_s(:,C) \bullet (C-1)\right],$$
$$\mathbf{U}_t = \frac{1}{n_t} \left[\widehat{\mathbf{Y}}_t(:,1:C)_{\hat{c}\neq 1}, \ldots, \widehat{\mathbf{Y}}_t(:,1:C)_{\hat{c}\neq C}\right],$$

where, $\mathbf{Y}_s$ and $\widehat{\mathbf{Y}}_t$ denote the $c-th$ column of $\mathbf{Y}_s$ and $\widehat{\mathbf{Y}}_t$, respectively. $\mathbf{Y}_s(:,c) \bullet (C-1)$ denotes that $\mathbf{Y}_s(:,c)$ is repeated $C-1$ times. The Eq. (7) can be further expressed to matrix form as the follows:

$$\mathcal{M}_d\left(\mathbf{X}_s, \mathbf{X}_t, \mathbf{A}\right) = \mathrm{tr}(\mathbf{A}^\top \mathbf{X} \mathbf{U} \mathbf{X}^\top \mathbf{A}), \tag{8}$$

where,

$$\mathbf{U} = \left[\begin{array}{cc} \mathbf{U}_s \mathbf{U}_s^\top & -\mathbf{U}_s \mathbf{U}_t^\top \\ -\mathbf{U}_t \mathbf{U}_s^\top & \mathbf{U}_t \mathbf{U}_t^\top \end{array} \right]. \tag{9}$$

In this article, $\mathcal{M}_t$ and $\mathcal{M}_s$ are integrated and defined as discriminant distribution distance as

$$D(\mathbf{X}_s, \mathbf{X}_t) = \mathcal{M}_t\left(\mathbf{X}_s, \mathbf{X}_t, \mathbf{A}\right) - \mu \mathcal{M}_d\left(\mathbf{X}_s, \mathbf{K}_t, \mathbf{A}\right) = \mathrm{tr}\left(\mathbf{A}^\top \mathbf{X} \left(\sum_{c=0}^{C} \mathbf{M}_c - \mu \mathbf{U}\right) \mathbf{X}^\top \mathbf{A}\right). \tag{10}$$

Discriminant distribution alignment strategy is used to enhance perdition performance by Eq. (10) can to improve the transferability and discriminability.

E. Local Structure Preserving

From a geometric perspective, if two instances in intra-class are close in the intrinsic geometry of the data distribution, then their transforming should also be close [46]. However, the transformed features cannot well preserve original local relationship consistency. For example, the distance between two instances in intra-class but from the different projects may be expanded after feature transformation, resulting in this distance being longer than the distance between two instances in different classes but from the same projects. In an attempt to solve these problems, we ensure that instances in intra-class stay close after feature transformation and keep instances in inter-class far from each other. To preserve the local structure in the transforming subspace, we have defined two embedding graphs (reward graph penalty graph)as expressed as below.

Reward Graph: $\mathbf{x}_i^c$ is connected with $\mathbf{x}_j^c$, where $\mathbf{x}_j^c$ is one of the k nearest neighbors of $\mathbf{x}_i^c$, $\mathbf{x}_i^c \in \mathbf{X}^c$ and $\mathbf{x}_j^c \in \mathbf{X}^c$.

Penalty Graph: $\mathbf{x}_i^c$ is connected with $\mathbf{x}_j^{\widehat{c}}$, where $\mathbf{x}_j^{\widehat{c}}$ is one of the k nearest neighbors of $\mathbf{x}_i^c$, $\mathbf{x}_i^c \in \mathbf{X}^c$ and $\mathbf{x}_j^{\widehat{c}} \notin \mathbf{X}^c$ $(c \neq \widehat{c})$.

The connection edges between the nodes of the two graph are assigned weights

$$\mathbf{W}_{ij} = \begin{cases} \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2}\right), & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are conencted} \\ 0, & \text{otherwise.} \end{cases}$$

We denote the pre-defined reward and penalty weight matrices as $\mathbf{W}^r$ and $\mathbf{W}^p$, respectively. The intra-class and inter-class scatter matrices are respectively defined as follows ($\mathbf{S}_b$ denotes the inter-class scatter matrix, and $\mathbf{S}_w$ denotes the intra-class scatter matrix):

$$\mathbf{S}_w = \frac{1}{2} \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} W_{ij}^r \left\| \mathbf{x}_i^c - \mathbf{x}_j^c \right\|^2 = \mathrm{tr}(\mathbf{A}^\top \mathbf{X} \mathbf{L}^r \mathbf{X}^\top \mathbf{A}), \tag{11}$$

$$\mathbf{S}_b = \frac{1}{2} \sum_{i=1}^{n_c} \sum_{j=1}^{n-n_c} W_{ij}^p \left\| \mathbf{x}_i^c - \mathbf{x}_j^{\widehat{c}} \right\|^2 = \mathrm{tr}(\mathbf{A}^\top \mathbf{X} \mathbf{L}^p \mathbf{X}^\top \mathbf{A}), \tag{12}$$

where, $\mathbf{L}^r$ and $\mathbf{L}^p$ respectively represent the reward Laplacian graph and the penalty Laplacian graph. ($\mathbf{L}^r = \mathbf{D}^r - \mathbf{W}^r$, $\mathbf{L}^p = \mathbf{D}^p - \mathbf{W}^p$). $\mathbf{D}_r$ and $\mathbf{D}_p$ are diagonal matrices whose each diagonal entry is $D_{ii}^r = \sum_j W_{ij}^r$ and $D_{ii}^p = \sum_j W_{ij}^p$, respectively. The geometric structure regularization can be expressed as follows:

$$\mathcal{G}\left(\mathbf{X}_s, \mathbf{X}_t, \mathbf{A}\right) = \mathbf{S}_w - \mathbf{S}_b = \mathrm{tr}(\mathbf{A}^\top \mathbf{X} (\mathbf{L}^r - \mathbf{L}^p) \mathbf{X}^\top \mathbf{A}). \tag{13}$$

F. Subspace Learning

Finally, the objective function in Eq. (1) can be reformulated byto maximize preserve the geometric structure of project instances.

$$\underset{\mathbf{A},\Phi}{\mathcal{L}} = \min \operatorname{tr}\left(\mathbf{A}^\top \mathbf{X}\left(\sum_{c=0}^{C} \mathbf{M}_c - \mu \mathbf{U} + \eta(\mathbf{L}^r - \mathbf{L}^p)\right)\mathbf{X}^\top \mathbf{A}\right) + \lambda \|\mathbf{A}\|_F^2, \tag{14}$$

$$\text{s.t.} \quad \mathbf{A}^\top \mathbf{X}\mathbf{H}\mathbf{X}^\top \mathbf{A} = \mathbf{I},$$

where, $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}$ is the centering matrix to avoid trivial solutions, and $\mathbf{I} \in \mathbb{R}^{n\times n}$ is the identity matrix. According to the constrained optimization, deriving the Lagrange function finds solution of problem Eq. (14). The Lagrange function is:

$$\underset{\mathbf{A},\Phi}{\mathcal{L}} = \min \quad \operatorname{tr}\left(\mathbf{A}^\top\left(\mathbf{X}\left(\sum_{c=0}^{C} \mathbf{M}_c - \mu \mathbf{U} + \eta(\mathbf{L}^r - \mathbf{L}^p)\right)\mathbf{X}^\top + \lambda \mathbf{I}\right)\mathbf{A}\right) + \operatorname{tr}\left(\left(\mathbf{I} - \mathbf{A}^\top \mathbf{X}\mathbf{H}\mathbf{X}^\top \mathbf{A}\right)\Phi\right), \tag{15}$$

where, $\Phi = diag(\phi_1, ..., \phi_d)$ is a diagonal matrix with the Lagrange multipliers. By setting the derivative of Eq. (15) $\frac{\partial \mathcal{L}}{\partial \mathbf{A}} = 0$, the optimization can be solved as a eigen-decomposition problem displayed as follows:

$$\left(\mathbf{X}\left(\sum_{c=0}^{C} \mathbf{M}_c - \mu \mathbf{U} + \eta(\mathbf{L}^r - \mathbf{L}^p)\right)\mathbf{X}^\top + \lambda \mathbf{I}\right)\mathbf{A} = \mathbf{X}\mathbf{H}\mathbf{X}^\top \mathbf{A}\Phi. \tag{16}$$

By taking the first d_0 smallest eigenvectors, the optimal solution of transformation matrix $\mathbf{A}$ is obtained. We can acquire the new representation $\mathbf{Z}_s = \mathbf{A}^\top \mathbf{X}_s$ and $\mathbf{Z}_t = \mathbf{A}^\top \mathbf{X}_t$. In summary, Algorithm 1 presents the pseudo code of the proposed LPDA approach.

Algorithm 1 Locality Preserving and Distribution Alignment (LPDA)

```
Input: Labeled source project: $\mathcal{D}_s = \{\mathbf{X}_s, \mathbf{Y}_s\}$; Unlabeled target project: $\mathcal{D}_t = \{\mathbf{X}_t\}$; Subspace dimension $d_0$; The number of iterations $T$; Number of nearest neighbors $k$; The trade-off hyper-parameters $\mu$, $\eta$ and $\lambda$.
Initialize pseudo labels $\widehat{\mathbf{Y}}_t$ of target project by using the classification model trained the source project.
Let $\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_t]$;
for $z$=1:$T$ do
    Construct MDD matrices $\mathbf{M}_0$ and $\mathbf{M}_c$ by Eq. (3) and Eq. (5);
    Construct discriminability matrix $\mathbf{U}$ by Eq. (9);
    Construct the reward Laplacian graph $\mathbf{L}^r$ and penalty Laplacian graph: $\mathbf{L}^p$;
    Solve Eq. (16) and take the $d_0$ smallest eigenvectors to construct $\mathbf{A}$;
    Obtain the source feature presentations $\mathbf{Z}_s = \mathbf{A}^\top \mathbf{X}_s$;
    Obtain the source feature presentations $\mathbf{Z}_t = \mathbf{A}^\top \mathbf{X}_t$;
    Train a classifier $f(\cdot)$ by using the source features presentations $\mathbf{Z}_s$ and labels $\mathbf{Y}_s$;
    Update the target pseudo labels $\widehat{\mathbf{Y}}_t$ by using the classifier $f(\cdot)$.
Output: The transformation matrix $\mathbf{A}$.
```

G. Computational Complexity

In this subsection, we present an analysis to time complexity of the proposed approach in Algorithm 1. The computational cost of solving eigen-decomposition problem is $O(T \times d_0 \times d^2)$, of constructing the MMD matrices, Laplacian graphs and the two discriminability matrix is $O(T \times C \times n^2)$, and of all other steps is $O(T \times d \times n)$. Thus, the overall computational complexity is $O(T \times d_0 \times d^2 + 3 \times T \times C \times n^2 + T \times d \ \times n)$.

TABLE I. ESSENTIAL INFORMATION OF THE SOFTWARE PROJECTS APPLIED IN THIS ARTICLE

Dataset	Project	# of metrics	# of total instances	% rate defective
Promise	ant-1.7	20	745	22.28
	ivyv-2.0	20	352	11.36
	jedit-4.1	20	312	25.32
	log4j-1.0	20	135	25.19
	lucene-2.2	20	247	58.30
	pio-2.0	20	314	11.78
	synapse-1.1	20	222	27.03
	tomcat	20	858	8.97
	xerces-1.4	20	588	74.32
NASA	CM1	37	327	12.84
	MW1	37	253	10.67
	PC1	37	705	8.65
	PC3	37	1077	12.44
	PC4	37	1287	13.75
AEEEM	EQ	61	324	39.81
	JDT	61	997	20.66
	LC	61	691	9.26
	ML	61	1862	13.16
	PDE	61	1497	13.96
ReLink	Apache	26	194	50.52
	Safe	26	56	39.29
	ZXing	26	399	29.57

IV. EXPERIMENTAL SETUP

A. Benchmark Datasets

To assess the put forward method, a total of 22 publicly software projects from four different software repositories, including AEEEM [47], NASA [48], Relink [49] and PROMISE are applied in our experiments. Table I offers comprehensive information about these projects.

AEEEM includes five software project [47]. Software module (instance) granularity is classified in AEEEM. Each instance includes 61 metrics (features), among them five entropy-of-change metrics, 17 code metrics, five previous-defect metrics, 17 entropy-of-code metrics and 17 churn-of-code metrics.

NASA is the most popular software defect data in previous studies. Each project signifies a software system, encompassing static code metrics along with associated defect labels. The static code metrics encompass attributes like McCabe, Halstead, lines of code, among others, and are valuable for predicting software quality and defects. The static code metrics encompass McCabe complexity, Halstead intricacy, code line count, and similar factors. These measurements offer valuable insights into software quality and predicting defects. In the research, we selected five projects that shared common feature spaces and merged them to create 20 cross-project predictive tasks. *Promise* is collected by Jureczko and Madeyski [50] and , comprises 20 class-level metrics, including CK metrics and QMOOD metrics. In the article, we selected 10 open projects and then used these projects to combine 90 cross-project prediction tasks.

ReLink is denoted by Wu et al. containing three open projects (i.e Apache, Safe and ZXing) They can be combined into cross-project prediction six tasks.

B. Performance Indicators

In this article, applying the four performance indicators, including F-measure, Balance, MCC and AUC evaluate our method and comparison methods. A software defect task typically yields one of four typical output results:

True Positive (TP): Correctly predicted defective instances.

True Negative (TN): Correctly predicted non-defective instances.

False Positive (FP): Incorrectly predicted defective instances.

False Negative (FN): Incorrectly predicted non-defective instances.

PD (aka. *recall or sensitivity*) determines the defective instances correctly predicted. The higher PD, the lower the cost generated by type II mis-classifications. *PF*(aka. *false positive rate*) is the ratio of non-defective instances that are incorrectly predicted.*Precision* is the ratio of correctly predicted instances that are defective They are denoted as the follows.

$$PD = \mathit{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}};$$
$$PF = \frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}};$$
$$Precision = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}.$$

F-measure is a harmonic mean of *Precision* and *Recall*, which is denoted as

$$\textit{F-measure} = \frac{(1+\theta^2)\times \mathit{Recall} \times \mathit{Precision}}{\mathit{Recall} + \theta^2 \times \mathit{Precision}}.$$

where, θerves as a bias parameter that determines the relative significance of *Recall* and *Precision*. There are three F-measure variants, i.e., $\mathbf{F_1}$ ($\theta = 1$) treats precision and recall equally, $\mathbf{F_{0.5}}$ ($\theta = 0.5$) prefers precision, and $\mathbf{F_2}$ ($\theta = 2$) prefers recall. There are two types of error in defect prediction. The first type of misclassification is the prediction of non-defect instances as defect instances. The second type of misclassification is when a defect instance is predicted to be a non-defect instance. The cost of the latter error is higher than that of the former error. Defect prediction mainly concerns finding as many defective instances as possible, which is consistent with the evaluation of performance measurement bias to recall. Thus, defect prediction more emphasize that Recall is more important than Precision(θ is set as 2).

Balance is the normalized Euclidean distance from the ideal point (1,0) to the actual point (*PD, PF*) in the ROC curve[51], [52], which is denoted as

$$\mathit{Balance} = 1 - \frac{\sqrt{(1-PD)^2 + (0-PF)^2}}{\sqrt{2}}.$$

MCC (Matthews Correlation Coefficient) is to measure the correlation coefficient between the actual and predicted outputs, which is denoted as:

$$MCC = \frac{\mathrm{TP}\times\mathrm{TN} - \mathrm{FP}\times\mathrm{FN}}{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}.$$

AUC(Area Under the Curve) The Area Under the Curve (AUC) refers to the area under the receiver operating characteristic curve (ROC). AUC is a comprehensive metric that effectively captures the trade-off and provides a better reflection of the overall performance of a prediction model.

C. Research Questions

To assess the predictive performance of the proposed LPDA, we delve into three research questions, in depth.

RQ1: Does LPDA perform better than the instances-based CPDP approaches?

As baselines for addressing this question, we employed five instance transfer approaches, which include ALL, NN-Filter, TNB, Peter-Filter, and DTB. Among these, TNB and DTB involve the reweighting of instances to mitigate the adverse influence of irrelevant cross-project data. In contrast, NN-Filter and Peter-Filter focus on filtering instances from the source project that are similar to the target project. These methods do not change the original feature space. Unlike these methods, the proposed method transfers the feature spaces while exploiting all the instances in the training step to avoid information loss. This research question is designed to investigate whether LPDA is superior to the instances-based CPDP methods in terms of CPDP performance improvement.

RQ2: Is LPDA more effective than the domain adaptation based CPDP methods?

We attempted to improve the CPDP performance from two aspects: the transferability between projects and the discriminability between class. Typically, domain adaptation-based CPDP methods investigate transferability by assuming that the source and target projects share a common distinguishing boundary. However, although the distribution gap between the two projects is narrowed after feature transformation, the instances from different classes are too close to be classified accurately near the decision boundary. The proposed LPDA method simultaneously explores the transferability and discriminability for CPDP tasks. This research question is designed to investigate whether the method considering both the transferability and discriminability is better able to improve the CPDP performance compared with other transfer learning methods.

RQ3: How do LPDA components have affect the prediction performance?

Since locality-preserving projection and distribution alignment are used in the proposed method, this research question is designed to investigate whether the components (i.e. transferable distribution alignment, discriminant distribution alignment and locality-preserving projection) can affect the prediction performance.

D. Statistical Testing

To better illustrate the effectiveness of the proposed method, we employ two statistical testing methods, namely the Wilcoxon signed-rank test and the Scott-Knott ESD test.

The Wilcoxon signed-rank test used to determine whether a significant difference exists between our method and each baseline for each cross-project task. Data distributions are identical without the assumption that they follow the normal distribution. Additionally, we employ the Win/Tie/Lose (W/T/L for brevity) evaluation to assess how many cross-project tasks our method can enhance in comparison to each baseline. Each entry's W/T/L implies that our method outperforms W cross-project tasks, ties on T cross-project tasks, and loses on L cross-project tasks.

The Scott-Knott ESD test is an expansion of the statistical methodology developed by Scott-Knott, which employs hierarchical cluster analysis to categorize a set of evaluation metrics into distinct, non-overlapping groups with statistically significant differences. This test consists of two stages: (1) the identification of a partition that maximizes the mean between different groups, and (2) either the separation into two distinct groups or the merging of any two groups with statistically significant differences and a negligible effect size into a single group. For a comprehensive explanation of the Scott-Knott ESD test, please refer to [53].

E. Experimental Settings

A total of 23 projects from four different repositories including AEEEM (5 projects), NASA (5 projects), ReLink (3 projects) and Promise (9 projects) are used in the paper. We we first identify all cross-project tasks in NASA, AEEEM, Relink and Promise. One project is selected as the target project, and the other projects from the same repository as the source. For example, when EQ is selected as the target project, the other projects separately are use as the source project. There are four cross-project tasks: JDT⇒EQ, LC ⇒EQ, ML⇒EQ, and PDE⇒ EQ. In total, there are 118 (9× 8+5×4 +5×4+3×2) cross-project tasks. We repeat each cross-project task 30 times and report the average values, each time we randomly select 90% of instances the source projects as the training set to train the model and all instances from the target project as test set, since a random selection of 90% of instances ensure that the training data are not consistently identical. In the article, Logistic Regression (LR) is selected as the foundational classifier. Due to its simplicity and effective performance in contrast to more intricate modeling methods, LR (Logistic Regression) has been a common choice in previous SDP research[18].

V. Experimental Results

A. Results for RQ1

To investigate the question, we apply some CPDP based-instance transferring methods as the baselines, including ALL, NN-Fifter, Petter-Fifter, TNB and DTB. ALL means that all the instances from the source project are used to train the prediction model without any instance filter and reweight process.

TABLE II. Average Values of Four Indicators for LPDA and Five Instance based CPDP Methods on the Different Dataset

Dataset	indicators	ALL	NN-Filter	Peter-Filter	TNB	DTB	LPDA
AEEEM	F-measure	0.404	0.423	0.409	0.532	0.486	**0.542**
	Balance	0.572	0.600	0.578	0.601	0.597	**0.696**
	MCC	0.204	0.233	0.239	0.248	0.276	**0.378**
	AUC	0.656	0.662	0.611	0.716	0.669	**0.748**
NASA	F-measure	0.358	0.417	0.317	0.418	0.393	**0.474**
	Balance	0.603	0.633	0.536	0.606	0.603	**0.670**
	MCC	0.179	0.186	0.193	0.202	0.206	**0.233**
	AUC	0.628	0.686	0.563	0.709	0.662	**0.724**
Promise	F-measure	0.442	0.482	0.498	0.459	0.457	**0.563**
	Balance	0.596	0.624	0.639	0.658	0.662	**0.702**
	MCC	0.250	0.316	0.256	0.322	0.259	**0.338**
	AUC	0.662	0.683	0.690	0.722	0.731	**0.746**
RELINK	F-measure	0.543	0.534	0.569	0.587	0.546	**0.767**
	Balance	0.607	0.640	0.638	0.595	0.637	**0.708**
	MCC	0.262	0.292	0.269	0.289	0.280	**0.320**
	AUC	0.669	0.700	0.702	0.641	0.701	**0.745**

TABLE III. Wilcoxon Signed-rank Test Results of LPDA Against Each Instance based CPDP Method

Dataset	Against(W/T/L)					
	indicators	ALL	NN-Filter	Peter-Filter	TNB	DTB
AEEEM	F-measure	20/0/0	19/1/0	19/1/0	14/2/4	15/3/2
	Balance	20/0/0	18/2/0	18/2/0	16/0/4	13/4/3
	MCC	19/0/1	19/0/1	18/0/2	17/1/3	14/2/4
	AUC	18/1/1	17/3/0	20/0/0	13/5/2	16/2/2
NSNA	F-measure	18/1/1	14/3/3	18/1/1	15/1/4	14/3/3
	Balance	15/1/4	13/2/5	17/2/1	16/1/3	17/0/3
	MCC	16/1/3	17/0/3	16/1/3	14/1/5	14/1/5
	AUC	16/2/2	12/2/5	19/0/1	8/5/7	18/2/0
Promise	F-measure	61/3/8	55/6/11	46/4/22	47/4/22	56/4/12
	Balance	64/4/4	57/4/11	54/4/14	46/12/14	53/6/13
	MCC	52/5/15	51/15/6	49/6/17	43/20/9	50/4/18
	AUC	56/3/13	53/4/15	48/8/16	44/6/22	40/6/26
RELINK	F-measure	6/0/0	6/0/0	6/0/0	6/0/0	6/0/0
	Balance	6/0/0	6/0/0	6/0/0	5/1/0	5/1/0
	MCC	6/0/0	5/1/0	6/0/0	5/0/1	5/1/0
	AUC	6/0/0	4/2/0	6/0/0	6/0/0	4/2/0
Total	F-measure	105/7/ 6	94/13/11	89/9/20	81/16/21	91/13/14
	Balance	105/8/5	94/11/13	95/11/12	83/17/9	88/14/16
	MCC	93/9/16	92/19/7	89/10/19	79/25/14	83/11/24
	AUC	96/9/13	86/15/17	93/11/14	71/19/28	78/15/25

According to the data presented in Table II, the proposed LPDA consistently outperforms the five baseline methods across various indicators on all datasets. For instance, on the AEEEM dataset, LPDA achieves an average F-measure value of 0.542, which represents a substantial improvement, ranging from 1.87% (in comparison to TNB) to a remarkable 34.06% (in comparison to ALL), with an average enhancement of 21.6%. In terms of the average Balance score (0.696) achieved by LPDA, improvements range from 15.77% (in comparison to TNB) to 21.68% (in comparison to ALL), with an average boost of 18.01%. Moreover, the average MCC value (0.376) obtained with LPDA demonstrates significant improvements, varying from 36.81% (in comparison to DTB) to an impressive 85.59% (in comparison to ALL), with an average enhancement of 59.04%. The average AUC value (0.748) also shows positive trends, with improvements ranging from 4.47% (in comparison to TNB) to 22.38% (in comparison to NN-Filter), averaging a substantial 12.85% improvement when contrasted with the five instance-based CPDP methods. Concerning the 20 cross-project pairs on the AEEEM dataset, data from Table III reveal that LPDA exhibits a statistically significant superiority in at least 13 of these cross-project pairs across all indicators. Conversely, it may perform less favorably on most of the four cross-project pairs. Fig. 1 demonstrates that the median values of all four indicators by LPDA higher than the five baseline methods. In particular, the median F-measure, Balance and AUC of the Relink dataset by LPDA are similar or even better than to the maximum value achieved by the five baseline methods.

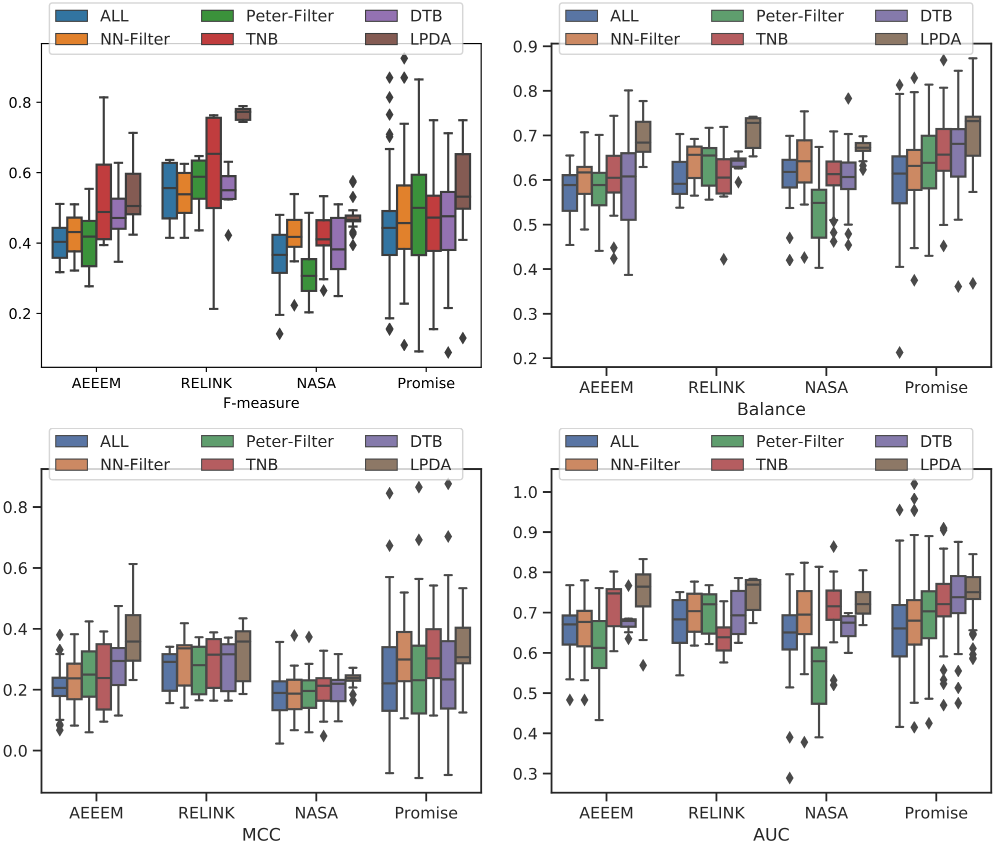


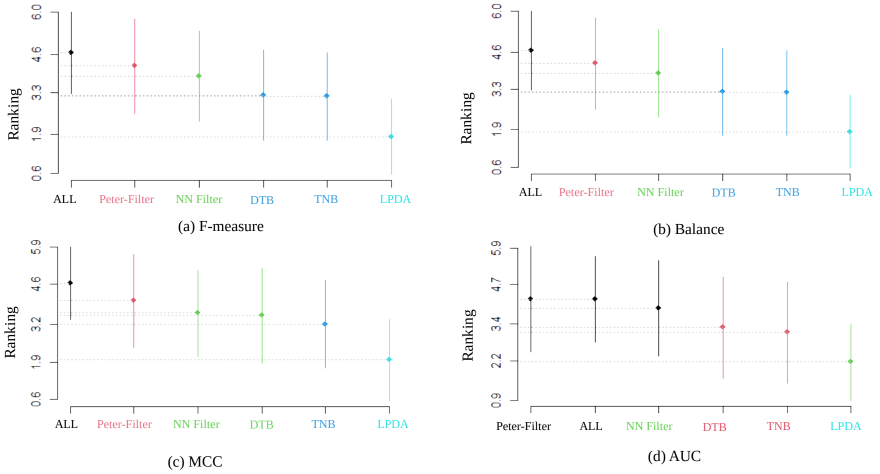
Fig. 1. Boxplots of f-measure, balance MCC, and AUC across all datasets for LPDA and the five instance-based CPDP methods.

On NASA dataset, the average F-measure value (0.474) by LPDA yields improvements between 13.40% (for TNB) and 49.45% (for Peter-Filter) with an average improvement of 25.92%, the average Balance value (0.670) by LPDA gains improvements between 5.77% (for NN Filter) and 24.89% (for Peter-Filter) with an average improvement of 12.65%, the average MCC value (0.233)by LPDA achieves improvements between 13.37% (for DTB) and 30.41%(for ALL) with an average improvement of 21.07%, and the average AUC value (0.724) gets improvements between 2.14% (for DTB) and 28.61% (for Pete-Filter) with an average improvement of 12.23% compared against the five instances-based CPDP methods. Three are also 20 cross-project pairs on NASA dataset. From Table III, the result shows that LPDA is significantly more accurate at least on 14 cross-project pairs and significantly less accurate at most on five cross-project pairs in terms of F-measure, Balance and MCC indicators. LPDA almost equal TNB (eight wins and seven losses) in terms of AUC. Compared to other the four baseline methods, LPDA is almost complete victory (at least 12 wins and at most 5 losses).

On the RELINE dataset, LPDA achieves an average F-measure value of 0.767, resulting in improvements ranging from 30.76% (compared to TNB) to 43.36% (compared to Peter-Filter), with an average improvement of 38.25%. The average Balance value, at 0.708 by LPDA, shows enhancements ranging from 10.64% (compared to NN-filter) to 18.98% (compared to TNB), with an average improvement of 13.66%. Furthermore, the average MCC value of 0.320 by LPDA demonstrates improvements ranging from 9.75% (compared to NN-filter) to 22.50% (compared to ALL), with an average improvement of 15.33%. The average AUC value of 0.745 also shows enhancements ranging from 6.16% (compared to Peter-Filter) to 16.16% (compared to DTB), with an average improvement of 9.27% when compared to the five instances-based CPDP methods. It is evident from Table III that LPDA outperforms the five baseline methods in terms of these four indicators, albeit it falls slightly short of them in certain aspects.

On Promise dataset, the average F-measure value (0.563) by LPDA yields improvements between 13.06% (for Peter-Filter) and 27.25% (for ALL) with an average improvement of 20.56%, the average Balance value (0.702) by LPDA gains improvements between 6.05% (for DTB) and 17.73% (for ALL) with an average improvement of 10.46%, the average MCC value (0.338) by LPDA achieves improvements between 4.84% (for TNB) and 34.90% (for ALL) with an average improvement of 21.84%, and the average AUC value (0.746) gets improvements between 2.08% (for DTB) and 12.72% (for ALL) with an average improvement of 7.07% compared against the five instances-based CPDP methods.

Table III shows the results of Wilcoxon signed-rank statistical test for each baseline method. By using the “W/T/L” evaluation, we can investigate 118 cross-project pairs in which LPDA can outperform other comparing method. As shown in the table, it is obvious that LPDA can achieve more positive



(a) F-measure

(b) Balance

(c) MCC

(d) AUC

Fig. 2. The results of scott-knott ESD test in f-measure, balance MCC, and AUC across all datasets for LPDA and other instance-based methods. The smaller ranking, the better performance.

results in terms of F-measure,Balance, MCC and AUC measure indicators when compared with other competing methods. For example, LPDA has significant superiorities to TNB on 81/118 (81 out of 118 cross-project pairs) in F-measure, and 83, 79 and 71 in Balance, MCC, and AUC, respectively. Fig. 2(a)-(d) present the results of Scott-Knott ESD test for the proposed LPDA and five baseline methods in terms of F-measure, Balance,MCC, and AUC, respectively. The x-axis and y-axis represent the method and ranking, respectively. The smaller the ranking, the better the performance. Each method corresponds to a bar denoting the range of ranking of this method on all cross-project pair tasks. The dot in the bar indicates the average ranking value. Different colors denote different groups with statistically significant differences. From these figures, we can see that LDPA obtains the smallest average ranking and thus is categorized into the top group which do not include any baseline method in terms of the four indicators.

B. Results for RQ2

In order to address this question, we have chosen five domain adaptation-based methods. A concise overview of these baseline methods is provided below:

1) *TCA*: This method aims to match marginal distribution between two projects [38] .
2) *TCA+*: An TCA variation improved in previous research [17] add customized normalization rules before distribution matching.
3) *JDM*: This method matches the marginal and conditional distributions simultaneously [15].
4) *BDA*: BDA [18] takes into account both the marginal and conditional distributions and dynamically assigns varying weights to them.
5) *TPTL*: TPTL [39] selects the benefits of source project selection and use transfer learning to construct a SDP model.

Table IV presents the mean values of the four metrics for both LPDA and the five domain adaptation-based methods across the four datasets. Fig. 3 illustrates box-plots representing the four metrics for all six methods across the entire set of datasets.

TABLE IV. Average Values of Four Indicators for LPDA and Five Domain Adaptation based CPDP Methods on the Different Dataset

Dataset	indicator	TCA	TCA+	JDM	BDA	TPTL	PLDA
AEEEM	F-measure	0.468	0.487	0.505	0.533	0.510	**0.542**
	Balance	0.671	0.645	0.674	0.689	0.647	**0.696**
	MCC	0.217	0.244	0.253	0.258	0.290	**0.378**
	AUC	0.701	0.716	0.725	0.703	0.729	**0.748**
NASA	F-measure	0.391	0.382	0.456	**0.484**	0.456	0.474
	Balance	0.651	0.572	0.646	0.666	0.660	**0.670**
	MCC	0.190	0.199	0.205	0.216	0.220	**0.233**
	AUC	0.657	0.675	0.705	0.707	**0.726**	0.724
Promise	F-measure	0.472	0.448	0.538	0.487	0.511	**0.563**
	Balance	0.656	0.678	0.693	0.699	**0.706**	0.702
	MCC	0.247	0.316	0.249	0.314	0.287	**0.338**
	AUC	0.668	0.677	0.693	0.690	0.705	**0.746**
RELINK	F-measure	0.636	0.639	0.632	0.643	0.703	**0.767**
	Balance	0.630	0.487	0.599	0.639	0.640	**0.708**
	MCC	0.271	0.306	0.282	0.306	0.295	**0.320**
	AUC	0.611	0.621	0.645	0.665	0.733	**0.745**

According to the data in Table IV, LPDA outperforms the five domain adaptation-based methods in all metrics when it comes to the AEEEM dataset, with higher average values.More specifically, LPDA compared with the five domain adaptation based methods achieves improvements of 1.55%–15.75% in F-measure , 0.92%–7.8% in Balance, 30.40%–74.42% in MCC, 2.26%–6.68% in AUC.

On the NASA dataset, the proposed LPDA achieves the best average values in terms of Balance and MCC, while BDA and TPTL achieve the best average values in terms of F-measure and AUC, respectively. More specifically, compared with the five baseline methods, LPDA achieves improvements ranging from 0.58% to 17.11% on Balance, and 5.92% to 22.99% on MCC, respectively. However, the average F-measure and AUC of LPDA are 1.95% and 0.2% lower than TNB and DTB, respectively. The results presented in Table IV show that average values in four evaluation indicators by

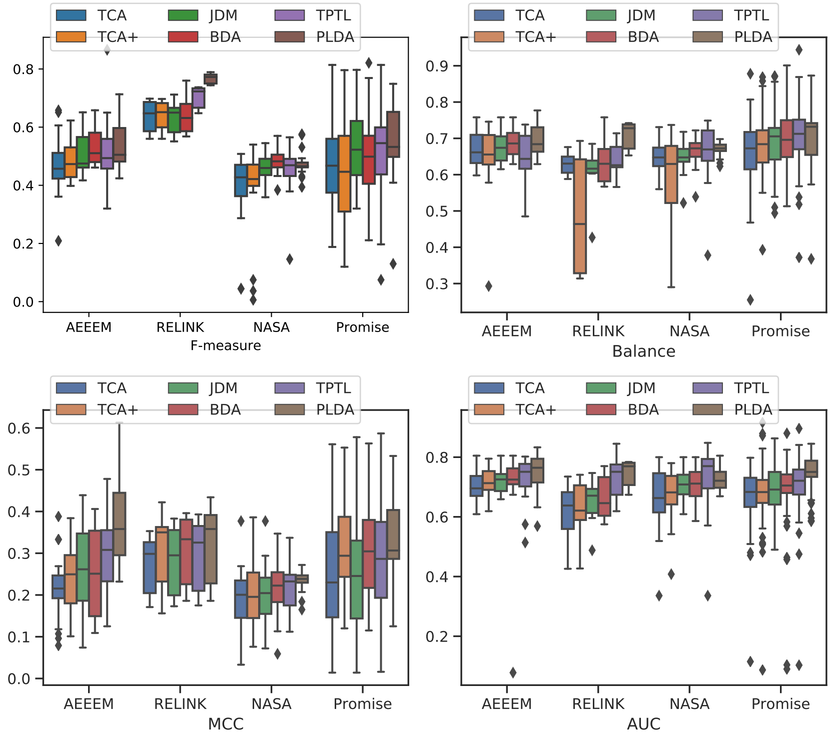


Fig. 3. Boxplots of f-measure, balance MCC, and AUC across all datasets for LPDA and the five domain adaptation based CPDP methods.

TABLE V. Wilcoxon Signed-rank Test Results of LPDA Against Each Domain Adaptation based CPDP Method

Dataset	Against(W/T/L)					
	indicators	TCA	TCA+	JDM	BDA	TPTL
AEEEM	F-measure	13/4/3	12/3/5	10/2/8	11/2/7	8/3/9
	Balance	14/2/4	14/3/3	13/3/4	8/9/3	14/3/3
	MCC	19/0/1	19/0/1	17/1/2	17/1/2	13/2/5
	AUC	18/1/1	14/4/2	13/4/3	11/4/5	13/4/3
NSNA	F-measure	13/4/3	12/3/5	10/2/8	11/2/7	8/3/9
	Balance	11/5/4	13/2/5	12/4/4	8/6/6	8/3/9
	MCC	14/3/3	14/3/3	13/4/3	11/4/5	9/5/6
	AUC	12/5/3	14/1/5	11/5/4	10/6/4	13/1/6
Promise	F-measure	54/1/17	50/1/21	39/2/31	49/7/16	44/3/25
	Balance	48/10/14	47/4/21	40/6/26	32/9/31	35/7/30
	MCC	55/2/15	45/15/12	52/4/16	38/19/15	44/5/23
	AUC	59/2/11	51/5/16	52/4/16	49/2/21	45/7/20
RELINK	F-measure	5/1/0	4/0/2	5/1/0	3/2/1	3/2/1
	Balance	6/0/0	6/0/0	6/0/0	5/1/0	5/1/0
	MCC	6/0/0	5/1/0	6/0/0	5/0/1	5/1/0
	AUC	6/0/0	6/0/0	6/0/0	6/0/0	3/2/1
Total	F-measure	90/10/18	84/8/26	71/8/39	71/18/29	72/13/33
	Balance	79/20/19	80/12/26	71/16/31	53/27/38	63/16/39
	MCC	93/9/16	82/21/15	87/13/18	69/29/20	69/17/32
	AUC	95/11/12	85/13/20	82/16/20	76/15/27	67/17/34

the proposed LPDA are the best average values on the Relink dataset. More specifically, compared with the five baseline methods, average value by LPDA gains the improvement of 9.17%-21.34% in terms of F-measure, of 10.53%–45.47% in terms Balance, of 4.61%–18.04% in terms MCC, and of 1.61%–21.82% in terms AUC. On the Promise dataset, the proposed LPDA achieves the best average values in terms of F-measure, MCC and AUC, while TPTL achieves the best average values in terms of Balance. To be specific, compared with the five baseline methods, LPDA achieves improvements ranging from 4.69% to 25.56% on F-measure, 6.97% to 36.79% on MCC, and 5.83% to 11.69% on AUC, respectively. However, the average value by LPDA is 0.58% lower than the best average value (for TPTL) among the five baseline methods in terms of Balance. From Table V, It is obvious that LPDA has more than 53 wins over Wilcoxon signed-rank test in any evaluation indicator. Taking TCA+ as an example, the "W/T/L" results show that LDPA has statistically significant improvements of 84/118 (84 out of 118 cross-project pair prediction), 80/118, 82/118 and 85/118 in F-measure, Balance MCC and AUC, respectively.

Fig. 3 depicts the boxplots of four indicators for the six methods on four datasets. This figure illustrates that the median values of all four indicators achieved by LPDA surpass those obtained by the five baseline methods on both the AEEEM and Relink datasets. On the NASA dataset, the median values of all four indicators by LPDA are not superior to these by the TPTL. On the Promise dataset, the values of Balance indicators

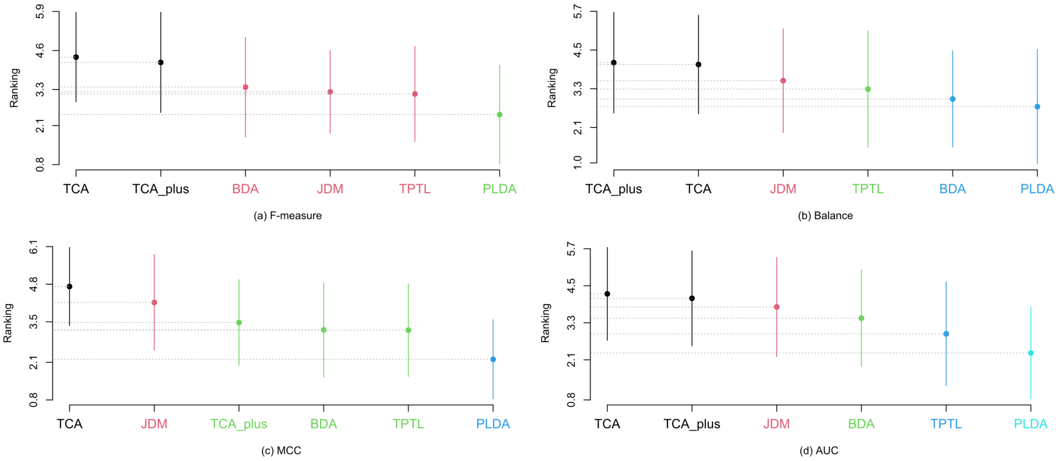


Fig. 4. The results of scott-knott ESD test in f-measure, balance MCC, and AUC across all datasets for LPDA and the five domain adaptation based CPDP methods.

by LPDA are slightly weaker than that by the TPTL.

Moreover, in Fig. 4(a)-(c), the outcomes of the Scott–Knott ESD test are presented for LPDA and the five domain adaptation-based methods across 118 cross-project pairs, considering F-measure, Balance, MCC, and AUC. The figures reveal that LPDA consistently achieves the lowest average ranking and is clearly separated into a distinct group concerning the four evaluation metrics. This indicates that LPDA significantly outperforms domain adaptation-based CPDP methods.

C. Results for RQ3

TABLE VI. Average Values of Four Indicators for LPDA and It's Variants on the Different Dataset

Dataset	indicator	LPDA_TA	LPDA_DA	LPDA_noDA	LPDA_noLP	PLDA
AEEEM	F-measure	0.474	0.498	0.496	0.504	**0.542**
	Balance	0.639	0.667	0.668	0.677	**0.696**
	MCC	0.336	0.363	0.358	0.361	**0.378**
	AUC	0.687	0.727	0.725	0.728	**0.748**
NASA	F-measure	0.389	0.399	0.392	0.413	**0.474**
	Balance	0.608	0.647	0.639	0.654	**0.670**
	MCC	0.214	0.212	0.212	0.228	**0.233**
	AUC	0.707	0.707	0.706	0.708	**0.724**
RELINK	F-measure	0.527	0.546	0.538	0.564	**0.767**
	Balance	0.687	0.670	0.701	0.666	**0.708**
	MCC	0.291	0.310	0.313	0.306	**0.320**
	AUC	0.713	0.717	0.717	0.732	**0.745**
Promise	F-measure	0.419	0.432	0.437	0.456	**0.563**
	Balance	0.662	0.666	0.683	0.684	**0.702**
	MCC	0.320	0.324	0.323	0.330	**0.338**
	AUC	0.719	0.728	0.727	0.729	**0.746**

LPDA has three key components, including locality preserving, transferable distribution transferable distribution alignment and discriminant distribution alignment. To investigate whether the proposed LPDA approach is more effective than other combinations of these three components, we specially conduct experiments to study the design of these components. We separately conduct LPDA_TA with only transferable distribution alignment (without locality preserving and discriminant distribution alignment), which is referred to as LPDA_TA, LPDA without discriminant distribution alignment, which is referred to as LPDA_noDA and LPDA without locality preserving, which is referred to as LPDA_noLP and LPDA with only discriminant distribution alignment which is referred to as LPDA_DA.

TABLE VII. Wilcoxon Signed-rank Test Results of LPDA Against Each Variant Method

Dataset	Against(W/T/L)				
	indicators	LPDA_TA	LPDA_DA	LPDA_noDA	LPDA_noLP
AEEEM	F-measure	17/1/2	14/2/4	16/1/3	16/1/3
	Balance	13/2/5	14/5/1	15/4/1	12/4/4
	MCC	13/3/4	10/8/2	14/6/0	12/5/3
	AUC	14/1/5	15/0/5	11/6/3	9/4/7
NSNA	F-measure	17/1/2	16/3/1	17/1/2	14/1/5
	Balance	16/2/2	11/3/6	16/2/2	10/6/4
	MCC	13/5/2	14/5/1	17/2/1	10/7/3
	AUC	11/4/5	12/3/5	12/4/4	12/3/5
Promise	F-measure	56/5/11	58/3/11	55/3/11	50/8/14
	Balance	48/10/14	47/4/21	40/6/26	32/9/31
	MCC	47/8/17	42/23/7	41/28/3	35/32/5
	AUC	45/16/11	40/13/19	43/15/14	43/16/13
RELINK	F-measure	6/0/0	6/0/0	6/0/0	5/1/0
	Balance	4/1/1	6/0/0	4/0/2	6/0/0
	MCC	4/1/1	3/2/1	2/3/1	5/1/0
	AUC	5/0/1	4/1/1	5/0/1	4/1/1
Total	F-measure	96/10/15	94/11/16	94/9/18	85/14/22
	Balance	82/22/17	76/22/23	81/28/12	70/32/19
	MCC	77/29/15	69/41/11	74/42/5	62/48/11
	AUC	75/24/22	71/20/30	71/28/22	68/27/26

Table VI shows the average values of F-measure, Balance, MCC and AUC on four different datasets of LPDA_TA, LPDA_DA, LPDA_noDA, PLDA_noLS and LPDA on Relink dataset, AEEEM dataset, NASA dataset and Promise dataset. In these table, the best values in each row are in bold. Fig. 5 depicts the boxplots of four indicators for the five methods. Table VII shows the results of Wilcoxon signed-rank statistical test. We present the following findings in Table VI, Table VII and Fig. 5. The results on the AEEEM dataset show that the proposed LDPA performs better than its four variants. To be specific, the improvement of LPDA over LPDA_TA, LPDA_DA, LPDA_noDA and PLDA_noLS on average is at least by 8.83%, 4.13%, 4.21%, 2.72% and at most by 14.38%, 8.87%, 12.48%, 8.86% in terms of F-measure, Balance, MCC and AUC, respectively. The results of Wilcoxon signed-rank statistical test shows that LPDA achieves more than 12 improvements with reference to Balance and F-measure , more than 10 improvements with reference to MCC, and more than nine improvements with reference to AUC, respectively.

On NASA dataset, the improvement of LPDA over LPDA_TA, LPDA_DA, LPDA_noDA and PLDA_noLS on average is at least by 14.69%, 2.38%, 2.26% , 2.27% and at most by 21.95%, 10.19%, 10.18%, 2.56% with regard to F-measure, Balance, MCC, and AUC, respectively. The results

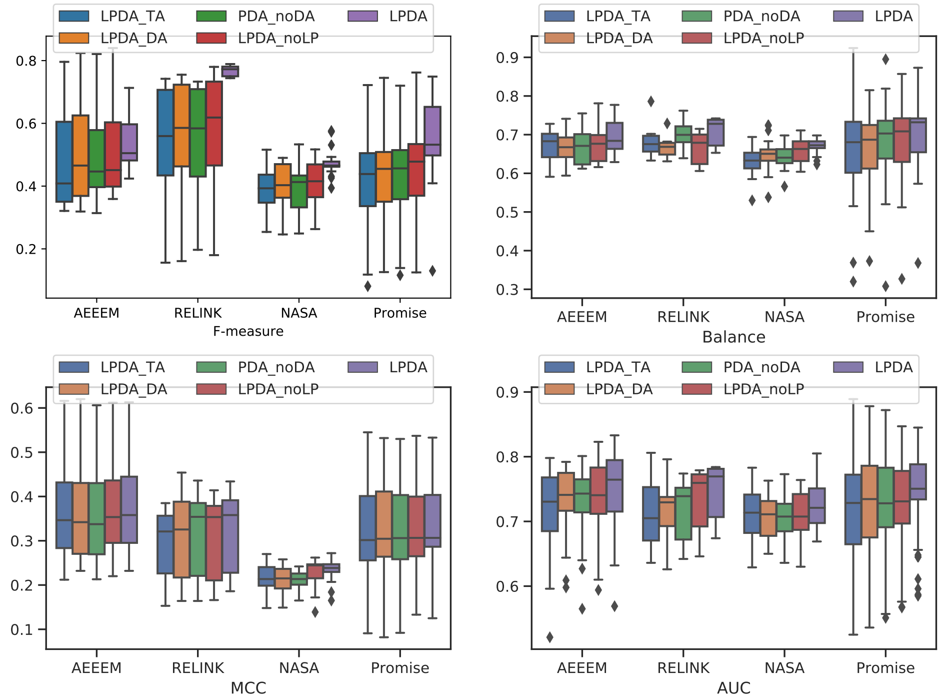


Fig. 5. Boxplots of f-measure, balance MCC, and AUC across all datasets for LPDA and it’s variant methods.

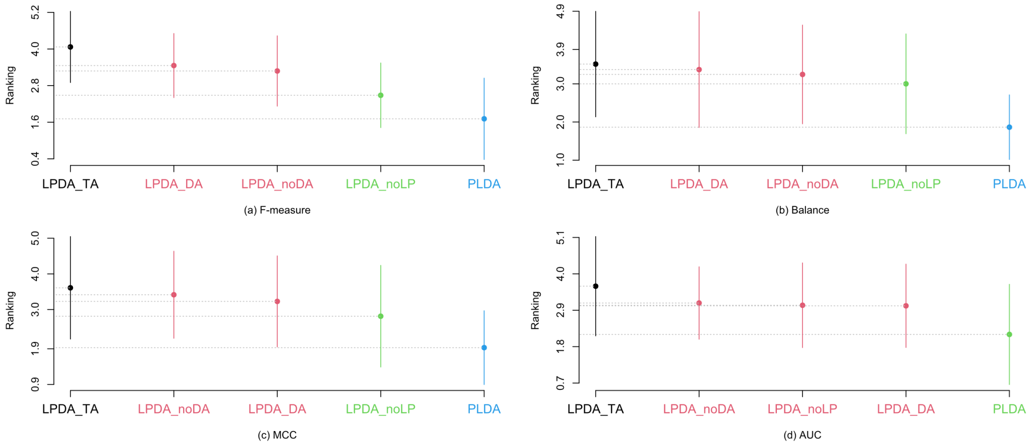


Fig. 6. The results of scott-knott ESD test in f-measure, balance MCC, and AUC across all datasets for LPDA and it’s variant methods. The smaller ranking, the better performance.

of Wilcoxon signed-rank statistical test shows that LPDA achieves more than 10 improvements with reference to Balance and MCC , more than 14 improvements with reference to F-measure, and more than 11 improvements with reference to AUC respectively.

On Relink dataset, the improvement of our method over LPDA_TA, LPDA_DA, LPDA_noDA and PLDA_noLS on average is at least by 36.11%, 1.04%, 2.33% , 1.76% and at most by 45.51%, 6.38%, 10.28%, 4.9% with regard to F-measure, Balance, MCC, and AUC, respectively. The results of Wilcoxon signed-rank statistical test shows that LPDA achieves more than four improvements with reference to Balance, AUC and F-measure. However, MCC of LPDA only wins on two cross-project pairs and ties on three cross-project pairs compared with LPDA_noDA.

On Promise dataset, the improvement of our method over LPDA_TA, LPDA_DA, LPDA_noDA and PLDA_noLS on average is at least by 23.44%, 2.55%, 2.28% , 3.32% and at most by 34.37%, 5.68%, 5.45%, 3.66% with regard to F-measure, Balance, MCC, and AUC, respectively. The results of Wilcoxon signed-rank statistical test shows that LPDA achieves more than 50 improvements with reference to F-measure, more than 32 improvements with reference to Balance, more than 35 improvements with reference to MCC and more than 43 improvements with reference to AUC respectively. As can be seen from the results shown in TableVII

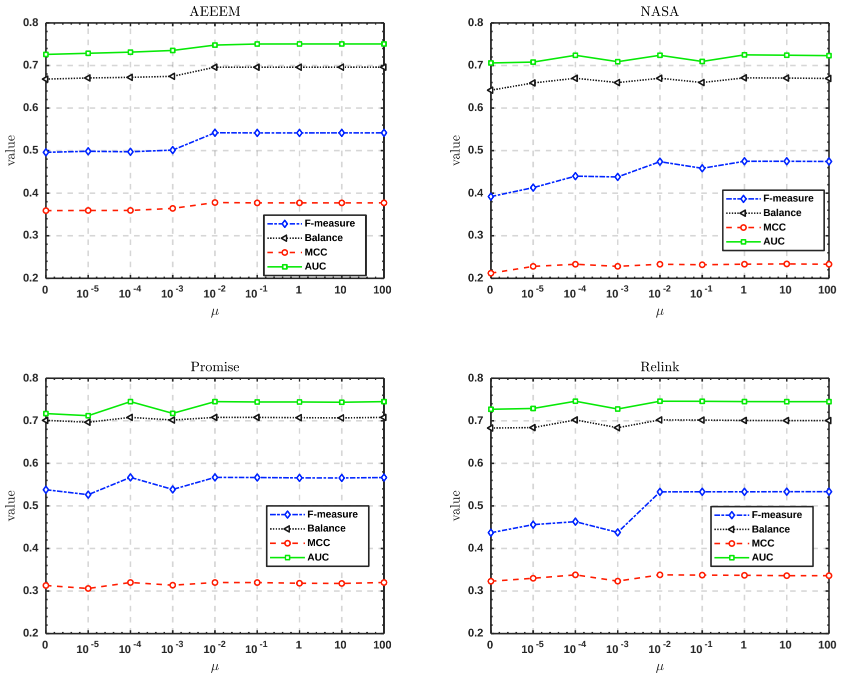


Fig. 7. Average f-measure, balance, MCC and AUC of LPDA on different datasets using differ trade-off parameter μ.

that LPDA attains more than 85 improvements in term of F-measure , more than 70 improvements with reference to Balance, more than 63 improvements with reference to MCC , and more than 68 improvements with reference to AUC, respectively.

VI. Discussion

A. Parameters Sensitivity

We explored the parameter sensitivity of LPDA to confirm that a diverse range of parameter values can be utilized to achieve a satisfactory level of performance. There are three parameters in LPDA, including the regularization parameter λ, the trade-off parameters μ and η. We change the values of these parameters in the range of $\{0, 0.00001, 0001, 0.001, 0.01, 0.1, 1, 10,1000\}$ to report the overall average F-measure, Balance, AUC, and MCC on different datasets at 30 random points of times. It is worth noticing when μ=0,LPDA becomes LPDA_noDA, which is without discriminant term, when η=0, LDPA becomes LDPA_noLP, which is without locality preserving term, and when μ=0 and η=0, LDPA becomes LPDA_TA.

Fig. 7 presents the aggregate average value of F-measure, Balance, MCC, and AUC for LPDA across various datasets, influenced by varying the trade-off parameter μ. This figure suggests that indicates that lower emphasis on discriminant distribution alignment (reflected by a smaller μ value) leads to less accurate LPDA predictions. Additionally, the g-measure exhibits steady fluctuation within the μ, parameter range of [0.01, 100]. We plotted the overall average F-measure, Balance, MCC and AUC of LPDA with different trade-off parameter η on four different datasets which are shown in Fig. 8. We find that LPDA is sensitive to η and robust to η in [0.001, 10].

We plotted the overall average F-measure, Balance, MCC and AUC of LPDA with different trade-off parameter λ on four different datasets which are shown in Fig. 9. We find LPDA is sensitive to λ and the robust results of λ on different datasets are different. We observe that $\lambda \in [0.001, 1]$ is the robust prediction result for AEEEM and NASA datasets. $\lambda \in [0.01, 1]$ and $\lambda \in [0.1, 100]$ are the robust parameter values for Promise and Relink datasets, respectively.

VII. Threats to Validity

A. Internal Validity

The primary challenges to internal validity stem from unmanaged variables affecting the experimental procedure. An instance of this is the emergence of defects during the reimplementation of baseline methods, which may occur during the coding phase. To mitigate these issues, especially for baselines with openly accessible source code (e.g., TCA and TPTL methods), we utilize the provided source code to minimize potential discrepancies. For baselines without accessible source code, we make every effort to ensure precise implementation by meticulously adhering to the instructions outlined in relevant research studies.

B. External Validity

External validity mainly concerns that our research can be generalized to other software projects. Twenty-two software projects are employed in our experiment. We evaluated our methods against ten CPDP approaches using five performance evaluation metrics. Additionally, we employed both the Wilcoxon signed-rank test and the Scott Knott-ESD test for further analysis. However, we still cannot guarantee the consistence of our method on other software projects not covered

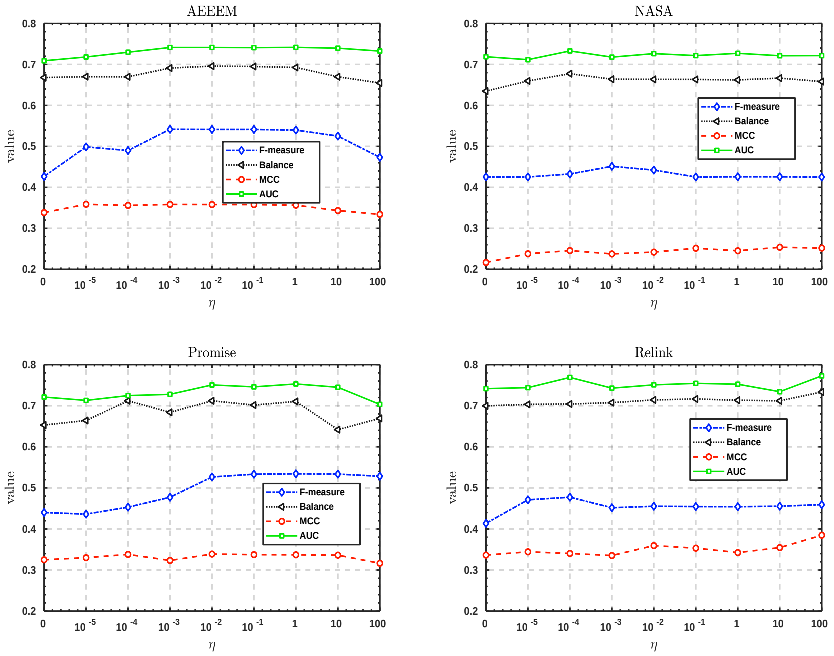


Fig. 8. Average f-measure, balance, MCC and AUC of LPDA on different datasets using differ trade-off parameter η.

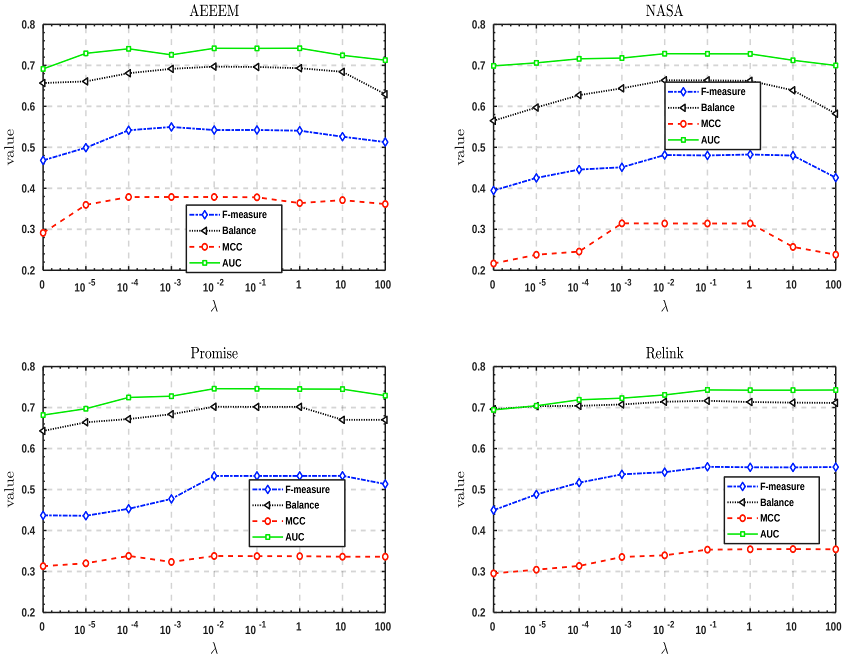


Fig. 9. Average f-measure, balance, MCC and AUC of LPDA on different datasets using differ trade-off parameter λ.

in this article. Further investigations on the applications of our research to commercial projects are needed.

VIII. Conclusion

In this paper, we propose a novel representation learning method LPDA for CPDP, which efficiently learned representations: 1) achieving transferability between two projects, and 2) preserving the local geometry to achieve discriminability between different categories. The local geometry is preserved by minimizing the distances between instances and their nearest neighbors within the same categories and maximizing the distances between instances and their neighbors in the different categories. The performance of LPDA is evaluated by five well-known evaluation indicators, including *Balance*, *F-measure*, *MCC* and *AUC*. We select ten state-of-the-art CPDP methods as baselines. We compare LPDA with ten baselines using the Wilcoxon signed-rank test and Scott-Knott ESD test (see Fig. 6). The experimental results on 22 software projects from four software repositories have shown that LPDA method is significantly superior to ten other CPDP methods. However, the proposed LPDA still has some limitations. LPDA does not consider important practical issues regarding the acquirement of useful knowledge from multiple external software projects and the application to projects. Moreover, LPDA method only considers traditional metric features and ignores effective high-level context features of source code. In our future work, we will further explore the potentials of LPDA.

Acknowledgment

The work was supported by the Guangdong Provincial Junior Innovative Talents Project for Regular Universities grant, No. 2023KQNCX005, and the Fundamental Research Funds for Central Universities.

References

[1] Li F, Yang P, Keung J W, et al. Revisiting ‘revisiting supervised methods for effort-aware cross-project defect prediction’[J]. IET Software, 2023, 17(4): 472-495.

[2] Chang R, Mu X, Zhang L. Software defect prediction using non-negative matrix factorization[J]. Journal of Software, 2011, 6(11): 2114-2120.

[3] Zou Q, Lu L, Qiu S, et al. Correlation feature and instance weights transfer learning for cross project software defect prediction[J]. IET Software, 2021, 15(1): 55-74.

[4] Thwin M M T, Quah T S. Application of neural networks for software quality prediction using object-oriented metrics[J]. Journal of systems and software, 2005, 76(2): 147-156.

[5] Zou Q, Lu L, Yang Z, et al. Joint feature representation learning and progressive distribution matching for cross-project defect prediction[J]. Information and Software Technology, 2021, 137: 106588.

[6] Al-Laham M, Kassaymeh S, Al-Betar M A, et al. An efficient convergence-boosted salp swarm optimizer-based artificial neural network for the development of software fault prediction models[J]. Computers and Electrical Engineering, 2023, 111: 108923.

[7] Wu F, Jing X Y, Sun Y, et al. Cross-project and within-project semisupervised software defect prediction: A unified approach[J]. IEEE Transactions on Reliability, 2018, 67(2): 581-597.

[8] Jing X Y, Wu F, Dong X, et al. An improved SDA based defect prediction framework for both within-project and cross-project class-imbalance problems[J]. IEEE Transactions on Software Engineering, 2016, 43(4): 321-339.

[9] Zain Z M, Sakri S, Ismail N H A. Application of Deep Learning in Software Defect Prediction: Systematic Literature Review and Meta-analysis[J]. Information and Software Technology, 2023: 107175.

[10] Kanwar S, Awasthi L K, Shrivastava V. Candidate project selection in cross project defect prediction using hybrid method[J]. Expert Systems with Applications, 2023, 218: 119625.

[11] Ren Y, Liu B, Wang S. Joint Instance and Feature Adaptation for Heterogeneous Defect Prediction[J]. IEEE Transactions on Reliability, 2023.

[12] Sharma U, Sadam R. How far does the predictive decision impact the software project? The cost, service time, and failure analysis from a cross-project defect prediction model[J]. Journal of Systems and Software, 2023, 195: 111522.

[13] Xia X, Lo D, Pan S J, et al. HYDRRA: Massively compositional model for cross-project defect prediction[J]. IEEE Transactions on software Engineering, 2016, 42(10): 977-998.

[14] Yu Q, Jiang S, Qian J. Which is more important for cross-project defect prediction: instance or feature?[C]//2016 International Conference on Software Analysis, Testing and Evolution (SATE). IEEE, 2016: 90-95.

[15] Qiu S, Lu L, Jiang S. Joint distribution matching model for distribution–adaptation-based cross-project defect prediction[J]. IET Software, 2019, 13(5): 393-402.

[16] Liu C, Yang D, Xia X, et al. A two-phase transfer learning model for cross-project defect prediction[J]. Information and Software Technology, 2019, 107: 125-136.

[17] Jiang W, Qiu S, Liang T, et al. Cross-project clone consistent-defect prediction via transfer-learning method[J]. Information Sciences, 2023, 635: 138-150.

[18] Xu Z, Pang S, Zhang T, et al. Cross project defect prediction via balanced distribution adaptation based transfer learning[J]. Journal of Computer Science and Technology, 2019, 34: 1039-1062.

[19] Ma Y, Luo G, Zeng X, et al. Transfer learning for cross-company software defect prediction[J]. Information and Software Technology, 2012, 54(3): 248-256.

[20] Bennin K E, Keung J, Phannachitta P, et al. Mahakil: Diversity based oversampling approach to alleviate the class imbalance issue in software defect prediction[J]. IEEE Transactions on Software Engineering, 2017, 44(6): 534-550.

[21] Feng S, Keung J, Yu X, et al. COSTE: Complexity-based OverSampling TEchnique to alleviate the class imbalance problem in software defect prediction[J]. Information and Software Technology, 2021, 129: 106432.

[22] Bennin K E, Tahir A, MacDonell S G, et al. An empirical study on the effectiveness of data resampling approaches for cross-project software defect prediction[J]. IET Software, 2022, 16(2): 185-199.

[23] Gong L, Jiang S, Bo L, et al. A novel class-imbalance learning approach for both within-project and cross-project defect prediction[J]. IEEE Transactions on Reliability, 2019, 69(1): 40-54.

[24] Su W C, Huang C Y. Application of Weighted Combinations of Activation Functions to Defect Prediction in Software Development[J]. IEEE Transactions on Reliability, 2023.

[25] Bhat N A, Farooq S U. An empirical evaluation of defect prediction approaches in within-project and cross-project context[J]. Software Quality Journal, 2023: 1-30.

[26] Wang R, Nie F, Hong R, et al. Fast and orthogonal locality preserving projections for dimensionality reduction[J]. IEEE Transactions on Image Processing, 2017, 26(10): 5019-5030.

[27] Wang Q, Breckon T P. Cross-domain structure preserving projection for heterogeneous domain adaptation[J]. Pattern Recognition, 2022, 123: 108362.

[28] Qiang W, Li J, Zheng C, et al. Robust local preserving and global aligning network for adversarial domain adaptation[J]. IEEE Transactions on Knowledge and Data Engineering, 2021.

[29] Liu J, Lei J, Liao Z, et al. Software defect prediction model based on improved twin support vector machines[J]. Soft Computing, 2023: 1-10.

[30] Zhou Y, Yang Y, Lu H, et al. How far we have progressed in the journey? an examination of cross-project defect prediction[J]. ACM Transactions on Software Engineering and Methodology (TOSEM), 2018, 27(1): 1-51.

[31] Zimmermann T, Nagappan N, Gall H, et al. Cross-project defect prediction: a large scale experiment on data vs. domain vs. process[C]//Proceedings of the 7th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering. 2009: 91-100.

[32] Li Z, Jing X Y, Wu F, et al. Cost-sensitive transfer kernel canonical correlation analysis for heterogeneous defect prediction[J]. Automated Software Engineering, 2018, 25: 201-245.

[33] Turhan B, Menzies T, Bener A B, et al. On the relative value of cross-company and within-company data for defect prediction[J]. Empirical Software Engineering, 2009, 14: 540-578.

[34] Peters F, Menzies T, Marcus A. Better cross company defect prediction[C]//2013 10th Working Conference on Mining Software Repositories (MSR). IEEE, 2013: 409-418.

[35] Kawata K, Amasaki S, Yokogawa T. Improving relevancy filter methods for cross-project defect prediction[C]//2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence. IEEE, 2015: 2-7.

[36] Bhat N A, Farooq S U. An improved method for training data selection for cross-project defect prediction[J]. Arabian Journal for Science and Engineering, 2022: 1-16.

[37] Hosseini S, Turhan B, Mäntylä M. A benchmark study on the effectiveness of search-based data selection and feature selection for cross project defect prediction[J]. Information and Software Technology, 2018, 95: 296-312.

[38] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE transactions on neural networks, 2010, 22(2): 199-210.

[39] Liu C, Yang D, Xia X, et al. A two-phase transfer learning model for cross-project defect prediction[J]. Information and Software Technology, 2019, 107: 125-136.

[40] Wang S, Liu T, Tan L. Automatically learning semantic features for defect prediction[C]//Proceedings of the 38th International Conference on Software Engineering. 2016: 297-308.

[41] Li J, He P, Zhu J, et al. Software defect prediction via convolutional neural network[C]//2017 IEEE international conference on software quality, reliability and security (QRS). IEEE, 2017: 318-328.

[42] Long M, Wang J, Ding G, et al. Transfer feature learning with joint distribution adaptation[C]//Proceedings of the IEEE international conference on computer vision. 2013: 2200-2207.

[43] Wang J, Chen Y, Hao S, et al. Balanced distribution adaptation for transfer learning[C]//2017 IEEE international conference on data mining (ICDM). IEEE, 2017: 1129-1134.

[44] Wang J, Feng W, Chen Y, et al. Visual domain adaptation with manifold embedded distribution alignment[C]//Proceedings of the 26th ACM international conference on Multimedia. 2018: 402-410.

[45] Zhang W, Wu D. Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation[C]//2020 international joint conference on neural networks (IJCNN). IEEE, 2020: 1-8.

[46] Wu H, Wu Q, Ng M K. Knowledge preserving and distribution alignment for heterogeneous domain adaptation[J]. ACM Transactions on Information Systems (TOIS), 2021, 40(1): 1-29.

[47] D'Ambros M, Lanza M, Robbes R. Evaluating defect prediction approaches: a benchmark and an extensive comparison[J]. Empirical Software Engineering, 2012, 17: 531-577.

[48] Siers M J, Islam M Z. Novel algorithms for cost-sensitive classification and knowledge discovery in class imbalanced datasets with an application to NASA software defects[J]. Information Sciences, 2018, 459: 53-70.

[49] Wu R, Zhang H, Kim S, et al. Relink: recovering links between bugs and changes[C]//Proceedings of the 19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering. 2011: 15-25.

[50] Jureczko M, Madeyski L. Towards identifying software project clusters with regard to defect prediction[C]//Proceedings of the 6th international conference on predictive models in software engineering. 2010: 1-10.

[51] Shao Y, Liu B, Wang S, et al. A novel software defect prediction based on atomic class-association rule mining[J]. Expert Systems with Applications, 2018, 114: 237-254.

[52] Shao Y, Liu B, Wang S, et al. Software defect prediction based on correlation weighted class association rule mining[J]. Knowledge-Based Systems, 2020, 196: 105742.

[53] Tantithamthavorn C, McIntosh S, Hassan A E, et al. An empirical comparison of model validation techniques for defect prediction models[J]. IEEE Transactions on Software Engineering, 2016, 43(1): 1-18.

Exploiting Deepfakes by Analyzing Temporal Feature Inconsistency

Junlin Gu, Yihan Xu, Juan Sun, Weiwei Liu
Jiangsu Vocational College of Electronics and Information,
China

Abstract—**In recent years, the rapid advancement of image generation technology has facilitated the creation of counterfeit images and videos, posing significant challenges for content authenticity verification. Malefactors can easily extract videos from social networks and generate their own deceptive renditions using state-of-the-art techniques. The latest Deepfake face forgery videos have reached an unprecedented level of sophistication, making it exceptionally difficult to discern signs of manipulation. While several methods have been proposed for detecting fraudulent media, they often target specific aspects, and as new attack methods emerge, these approaches tend to become obsolete. This paper presents a novel detection approach that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM). Initially, CNN is employed to extract image features from each frame of the input facial video, capturing subtle alterations and irregularities in manipulated content. Subsequently, the extracted feature sequence is used to train the LSTM network, mimicking the temporal consistency of human visual perception and enhancing the effectiveness of counterfeit video detection. To validate this methodology, a comprehensive evaluation is conducted using the FaceForensic++ dataset, affirming its proficiency in identifying Deepfake counterfeit videos.**

Keywords*—*Face forgery detection; Convolutional Neural Network; Long Short-Term Memory Network; time consistency

I. Introduction

The rapid and remarkable progress in machine learning technology has elevated the capabilities of video modification and production to an unprecedented level. A pivotal development in this arena is the widespread adoption of Generative Adversarial Networks (GANs), which have revolutionized the automatic generation of images and video synthesis through network model training [1]. Notably, the advent of Deepfake technology, a derivative of GANs, has enabled the seamless replacement of facial features in videos. After post-processing, these videos attain an exceptional degree of realism. However, this rapid technological advancement has brought forth a slew of significant societal challenges [2]. The proliferation of Deepfake technology raises concerns about privacy violations, and its misuse can potentially lead to legal liabilities. Despite diligent efforts by network oversight bodies, the digital landscape remains inundated with a vast volume of synthetic, manipulated videos. It is, therefore, imperative to expeditiously develop effective methods for detecting forged videos to address this burgeoning issue.

As Deepfake technology continues to evolve, the field of detection methods has made substantial progress. Researchers have delved deeply into deep learning models, including spatial domain methods [3], [4], [5] and temporal domain methods [6], [7], in a comprehensive effort to identify irregularities and inconsistencies inherent in Deepfake videos. These models autonomously extract and categorize features, thereby enhancing the accuracy of forged content detection. Moreover, the development and utilization of extensive datasets have played a pivotal role in advancing research on deep forgery detection. Datasets such as FaceForensics++ [8], Deeperforensics [9] have provided a wealth of real and fake video examples, serving as invaluable resources for researchers in this field. The adoption of multi-modal detection approaches, which combine visual data with audio, voice, and other sources of information, has significantly improved the precision and effectiveness of detection methods. Nevertheless, the field of deep forgery detection continues to grapple with an array of challenges. Adversaries consistently refine their Deepfake techniques, making detection increasingly intricate. Consequently, researchers are compelled to continually enhance detection methodologies to bolster their robustness and real-time performance, effectively responding to evolving forgery threats.

In alignment with these advancements, this paper introduces a temporal feature inconsistency analyzing method to enhance the accuracy of Deepfake forgery detection. Specifically, the proposed approach integrates a deep convolutional neural network for image feature extraction and incorporates an LSTM network to analyze correlations between feature sequences. Empirical findings substantiate the efficacy of this methodology, affirming its capacity to facilitate efficient and reliable deep forgery video detection. The main contributions of this paper are as follows:

1) We propose a deepfake detection framework based on the extraction of temporal inconsistencies, combining the feature extraction capabilities of CNN and the temporal feature analysis abilities of LSTM to achieve accurate detection of deepfakes.
2) We tested the algorithm's detection accuracy on video sequences of various lengths using the FaceForensics++ dataset. The experimental results indicate that our algorithm ensures both detection accuracy and computational efficiency when applied to video sequences of 40 frames in length.

This paper focuses on detecting deepfake videos using temporal continuity. Section II provides an overview of recent advancements in deepfake detection research. Section III details the proposed methodology, while Section IV validates its effectiveness through experiments measuring detection accuracy, computational efficiency, and related metrics. Finally, the Section V concludes the paper by summarizing our proposed scheme.

II. Related Work

Currently, research in the field of image forgery detection has made notable advancements. However, there is still a compelling need for further exploration, particularly in the context of real-world scenarios. The domain of forged video detection is predominantly categorized into two main classes: semantic detection and non-semantic detection. The specific detection methods within these categories are systematically organized and illustrated in Fig. 1.

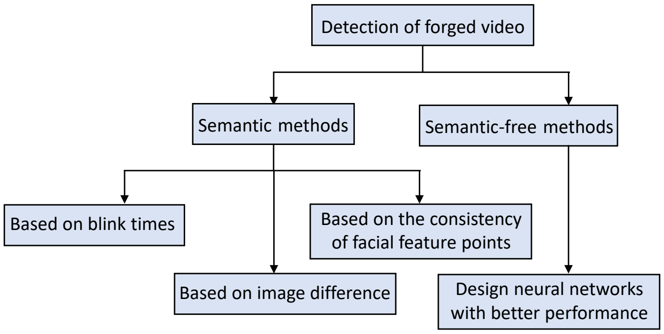

Detection of forged video
Semantic methods
Semantic-free methods
Based on blink times
Based on the consistency of facial feature points
Based on image difference
Design neural networks with better performance

Fig. 1. Classification of forgery video detection methods.

A. Semantic-based Detection Method

1) Classification using the number of blinks in the video: Blinking, denoting the swift opening and closing of eyelids, constitutes a notable behavioral trait. The generation of counterfeit videos through GAN models relies on extensive training data sourced from facial images. As a consequence, many genuine photos do not capture subjects with their eyes closed, leading to a distinctive lack of blink in the generated fake videos. From this vantage point, the absence of blinking emerges as a conspicuous discrepancy between counterfeit and authentic videos.

In the realm of computer vision, blink detection has garnered attention for diverse applications such as fatigue detection [10], [11], [12] and face spoofing detection [13]. Various approaches have been explored in this context. Sukno et al. [14] employed an active shape model in conjunction with optimal invariant features to delineate the eye contour, subsequently assessing eye state based on vertical eye displacement. Torricelli et al. [15] analyzed eye states through the comparison of consecutive frames. Divjak et al. [16] harnessed optical flow to capture eye motion, subsequently extracting the principal eye motion for analysis. Yang et al. [17] deployed parameterized parabolic curves to model eye shapes and fitted the model to individual frames for eyelid tracking.

Drutarovsky et al. [18] delved into the variance of vertical eye area movement as detected by the Viola-Jones algorithm. They further utilized a group of KLT trackers within the eye area, dividing each eye region into 3x3 subregions to calculate the average motion within each. Notably, the most recent development in forged video detection with a focus on blink motion is attributed to Li et al. [19]. Their approach discerns blink occurrences through a combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory Neural Networks (LSTM), ultimately rendering judgments regarding video authenticity based on blink frequencies.

However, this technology primarily hinges on the quantification of blink incidents. Crucially, GAN models employed for the generation of counterfeit videos are trained on a substantial corpus of facial images. In the event that malicious actors augment the training data with closed-eye facial images, the resultant Deepfake counterfeit videos will exhibit plausible blink occurrences, effectively undermining the blink-based detection mechanism.

2) Using the difference between the head pose of the person in the generated video and the head pose of the original video to classify: The face exchange algorithm is designed to generate faces of different individuals while preserving the original facial expressions. However, it is essential to note that the facial feature points of these two faces may not align. The positions of these feature points on the human face are intrinsically linked to crucial structures such as the eyes and mouth. Given that neural network synthesis algorithms cannot guarantee the exact replication of facial features between the original human face and the synthesized face, Yang et al. [20] introduced a novel approach to assess the head pose by comparing estimations derived from all facial feature points with those calculated solely from the central region.

This method is grounded in the observation of errors stemming from the integration of the synthesized face region into the original image. These errors become evident when attempting to estimate the three-dimensional head pose from the facial image. The authors empirically validated this phenomenon through a series of experiments and subsequently devised a classification method based on these observations. It's noteworthy, however, that this approach has yet to be evaluated using the latest Deepfake forged face datasets. Consequently, the question of whether it can effectively detect the most recent Deepfake videos remains an open challenge.

3) Classification by comparing the differences between the face area and the surrounding area: In the realm of image and video detection, recent strides have been taken towards identifying content generated by Generative Adversarial Networks (GANs). Notably, in the context of face exchange, where the original face image from one video is transposed onto the face image of another video, even after a series of fuzzy optimization processes, disparities inevitably emerge between the facial image and its surrounding context.

Li et al. [21] introduced a novel approach that leverages a Convolutional Neural Network (CNN) model to discern discrepancies between the facial region and its neighboring context, thereby facilitating the detection of forged faces. To simulate a broader spectrum of affine distorted faces across different resolutions, the authors trained four CNN models, namely VGG16 [22], ResNet50, ResNet101, and ResNet152 [23]. Subsequently, these models were evaluated by testing them on several synthetic videos sourced from YouTube, affirming the effectiveness of this method for detecting Deepfake forged videos.

B. Non-semantic Detection Method

In recent years, the field of digital image forensics has witnessed a notable integration of deep learning techniques. Rao and Ni [24] introduced a network dedicated to detecting image stitching, while Rahmouni et al. [25] demonstrated the capacity of deep learning to discriminate between computer-generated and photographic images. These developments underscore the robust performance of deep learning in the domain of digital forensics.

Indeed, traditional microscopic analysis relying on image noise becomes inapplicable within the constraints of compressed video environments, where image noise is often significantly denoised. Similarly, differentiating forged face images at a higher semantic level poses a considerable challenge for the human eye. To address these issues, Darius and Vincent et al. [26] proposed an intermediary approach, employing a deep neural network with a streamlined architecture for image detection. They presented two network structures tailored for detecting forged videos, achieving commendable detection results with a minimal computational overhead. Experimental results showcase an average detection accuracy of 98% for Deepfake counterfeit videos. To further validate the efficacy of this solution, considerable effort was devoted to visualizing the designed network layers and filters.

Another pioneering contribution in the realm of deepfake video detection was presented by Huy et al. [27], who harnessed capsule networks for detecting counterfeit videos across diverse scenarios. This work marked a significant advancement, as it was among the first to explore the application of capsule networks in the field of detection. Capsule networks, initially devised to address issues in digital forensics, were thoroughly examined. The authors conducted a comprehensive analysis and comparison against four mainstream datasets, affirming the superior performance of their method.

However, challenges still persist in the domain of Deepfake video detection, including issues related to the representativeness of datasets and the limited scope of detection. To address these concerns, this paper employs a combination of Convolutional Neural Networks for feature extraction and Long Short-Term Memory networks for analyzing temporal inconsistencies.

III. Problem Analysis and Proposed Method

A. Problems in Deepfake Generation Methods

This section briefly introduces the process of Deepfake generation, and analyzes the problems existing in Deepfake generation method according to its production process.

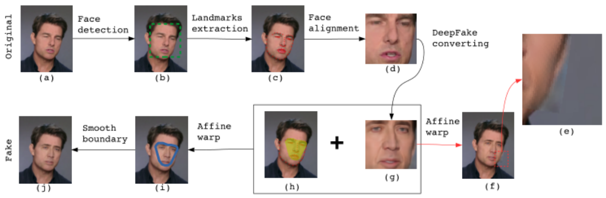


Fig. 2. Forged face generation process diagram.

As illustrated in Fig. 2, the process of generating a forged frame image within a video is detailed. Initially, the original image (a) undergoes face detection to delineate the facial region, depicted as the bounding box in (b). Subsequently, the facial feature points are extracted, and these extracted points are visualized in (c). These feature points convey essential facial characteristics, including facial orientation. Following necessary adjustments, the result in (d) is obtained, which then serves as input to a Generative Adversarial Network (GAN) to produce (g). The subsequent task involves seamlessly integrating (g) with the original image. Two distinct methods are employed for this integration. The first method entails directly replacing (g) with the original image (a) via an affine transformation, generating an image such as (f). However, it becomes evident from (e) that the replaced area does not seamlessly blend with the original image, resulting in noticeable discrepancies. The second approach involves initially identifying the region to be replaced based on the feature points detected in (c), as depicted in (h). This replacement region predominantly corresponds to the central area of the face. Subsequently, (g) is replaced in this region via an affine transformation. Finally, the boundary of the replacement region is softened and smoothed to enhance the image's overall realism. Totally, the problems of Deepfake generation technology are as follows:

1) Intra-frame: When introducing a new face into the target frame image, even with subsequent blurring and smoothing of the boundary, the central facial region tends to exhibit disparities in terms of color, brightness, and resolution when compared to the other areas within the target frame image.

2) Inter-frame: In the context of video face forgery, each image frame undergoes processing, and the GAN employed to generate facial images lacks the ability to retain knowledge of previous frames. In essence, the GAN lacks information about the facial content in the preceding frames, making it challenging to capture the temporal consistency between adjacent frames. Consequently, the facial expressions in consecutive frames may exhibit significant divergence, whereas genuine video sequences tend to maintain a higher degree of consistency in the facial expressions between adjacent frames.

B. The Proposed Method

Building upon David's approach [28], this paper leverages the incongruities present in intra-frame and inter-frame Deepfake video content for detection purposes. In addressing intra-frame disparities, Convolutional Neural Networks are harnessed to extract image features, with the objective of obtaining discriminative features capable of distinguishing genuine from fabricated videos. To address the issue of temporal continuity between frames, this paper adopts the Long Short-Term Memory network (LSTM) for detection. LSTM is a network well-suited for processing sequential data, allowing for the analysis and processing of features that exhibit temporal coherence.

Consequently, this experiment capitalizes on the inconsistency within the image content of Deepfake forged videos and the lack of continuity between adjacent frame images. The process commences with feature extraction performed on each frame within the video, and the resulting feature sequence is subsequently input into the LSTM network. The LSTM network is meticulously trained to identify Deepfake forged videos. The workflow of this solution is visually represented in Fig. 3.

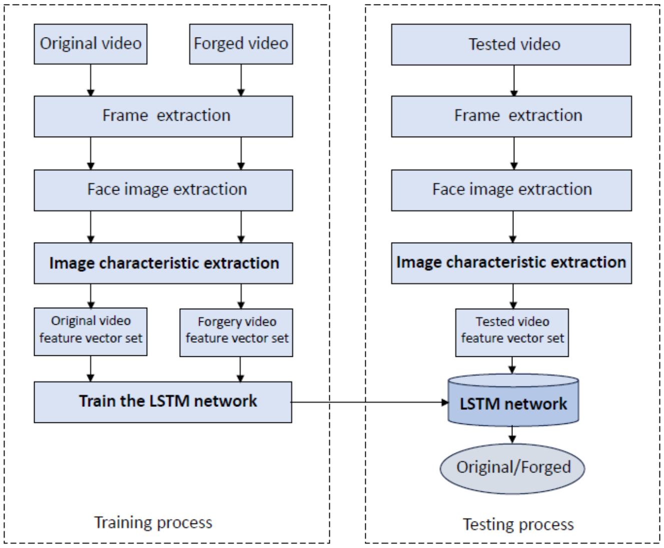


Fig. 3. The workflow of the solution.

1) Video frame extract: Before proceeding with spatial feature extraction, the initial step involves extracting individual frame images from the video footage. To comprehensively evaluate the minimum duration of a forged video that can be reliably and effectively detected, this experiment involves the selection of a sequence of consecutive N-frame images. The study investigates three distinct values for N: 20, 40, and 60 frames, providing a thorough examination of the impact of video length on detection accuracy.

2) Face image extraction: Prior to the spatial feature extraction phase, the fundamental process begins with isolating facial images from the video frames. This initial step holds significant importance for multiple reasons. Firstly, the facial region and its immediate surroundings inherently exhibit a higher degree of distinctiveness, rendering them a valuable asset in the task of differentiating genuine content from manipulated video segments. In our approach, facial detection techniques are employed to pinpoint and delineate prominent facial features within each frame. Once these facial feature points are successfully identified, the corresponding facial images are cropped and separated from the frames. This foundational step lays the groundwork for subsequent spatial feature extraction procedures, ensuring the preservation and utilization of the most informative and distinguishing component of the video - the human face - to enhance the robustness and effectiveness of subsequent detection processes.

3) Image characteristic extraction: Convolutional Neural Networks (CNN) have consistently proven their superiority in image classification tasks due to their exceptional feature extraction capabilities. In the context of forgery video detection, the process of face replacement inevitably introduces inconsistencies between the replaced face image and the surrounding context. These inconsistencies can be effectively captured by the image features extracted through CNN. This paper employs multiple CNNs to individually extract features from face images. The choice of the most suitable neural network model for forgery video detection is determined by comparing their classification performance. These CNNs are utilized to extract image features from video frames within the training and test datasets.

Considering the potential limitations of self-constructed network models in feature extraction, pre-trained network models on the ImageNet dataset, such as VGG19, Inception-V3, and ResNet, are also considered. After removing the last output layer, the output of the final fully connected layer serves as the feature representation for each frame image, thereby facilitating feature extraction from each frame.

4) LSTM network training: To analyzing the temporal inconsistency, the Long Short-Term Memory (LSTM) network is harnessed to analyze the feature sequence extracted by CNN. The LSTM network is equipped with a fully connected layer to map the features derived from LSTM into the ultimate forgery video detection probability. This training process culminates in the development of an LSTM network model designed to serve as a classifier for forged videos.

IV. Experiments

Aiming at the inconsistency in Deepfake forged video, this experiment uses Convolutional Neural Network to extract the spatial features of each frame image to obtain the continuous spatial features of the video, and then inputs the feature sequence into the LSTM network. The sequence features are extracted by the LSTM network to train the classifier, so as to realize the detection of forged video. This chapter introduces the experimental part.

A. Experiment Settings

1) Dataset: This study leverages the widely recognized FaceForensics++ dataset, an extension of the original FaceForensics dataset that has gained extensive adoption in the digital forensics community. The creation of this dataset involved a meticulous process. The dataset production team initiated the project by sourcing 1000 original video files from various online platforms. To ensure the suitability and quality of these videos for research purposes, a minimum resolution of 480p or higher was enforced for each selected video. In recognition of the potential confounding factor of facial occlusion, the production team embarked on a comprehensive manual segmentation process to painstakingly remove any occluded facial fragments within the videos. As a result, the dataset comprises a total of 1000 original videos, with each video yielding an extensive collection of 509,914 individual frames upon the extraction of each image frame. The primary focus of this research lies in the detection of Deepfake counterfeit videos. Therefore, the dataset predominantly draws upon the original video set available in the dataset, complemented by a dedicated Deepfake counterfeit video dataset. The construction and preparation of these datasets were carried out with exceptional care and precision, underscoring their pivotal role in facilitating robust and credible research in the domain of forged video detection. Within the dataset, the training set comprises 720 videos, while the validation and test sets collectively encompass 140 videos. The exact count of video frames included in each dataset can be found in Table I, providing a comprehensive overview of the dataset's composition. This rigorous dataset design and curation serve as an indispensable foundation, ensuring the availability of a diverse and comprehensive set of video data that is critical for advancing the field of forged video detection.

TABLE I. The Total Number of Video Frames Contained in Each Dataset

Dataset	Training set	Validation set	Testing set
Original	367,282	68,862	73,770
Face2Face	367,282	68,862	73,770
FaceSwap	292,376	54,630	59,672
Deepfakes	367,282	68,862	73,770

2) Experimental parameters: With the network architecture defined, the subsequent step involves data transmission to initiate the training phase. For this experiment, the Adam optimizer, acknowledged for its proficiency in optimizing deep learning models, is employed. The learning rate, a critical hyperparameter influencing convergence and training dynamics, is thoughtfully set to 1e-4 to promote a well-balanced learning process. During the training phase, a designated batch size of 8 examples is input into the network in each training iteration. This iterative process continues until the network achieves convergence, a pivotal juncture in the training cycle where the model has reached its optimal learning capacity. Subsequently, the trained network model undergoes rigorous testing to assess its performance, with a specific focus on detection accuracy. In Table II, provided below for reference, the experiment accounts for the variability in the number of nodes within the fully connected layers in distinct network configurations. To reconcile this variance, systematic adjustments are made to the parameters governing the Long Short-Term Memory (LSTM) layers and the subsequent fully connected layers. These parameter modifications are executed with precision, ensuring the seamless and coherent operation of the network across various architectural configurations.

B. Face Image Extraction

The fundamental premise of this algorithm centers on harnessing discrepancies within the facial region, recognizing that this region exhibits substantial variations indicative of manipulation. Nonetheless, it's essential to acknowledge that a complete video frame contains a plethora of information extending beyond the facial area, which may not be pertinent to the analysis at hand. Therefore, a critical preprocessing step involves the extraction of facial images, with a specific focus on precisely delineating the face and its immediate surroundings. This segmentation process is pivotal in isolating the region of interest, as depicted in Fig. 4, and subsequently refining the dataset for effective analysis. This strategic extraction not only mitigates computational overhead but also streamlines the subsequent processes of feature extraction and classification, thereby enhancing the efficiency and accuracy of Deepfake video detection.

TABLE II. Parameter Settings of the Networks

Network	VGG19	Inception-V3	ResNet50
LSTM	4096	2048	2048
Fully connected layer	512	512	512

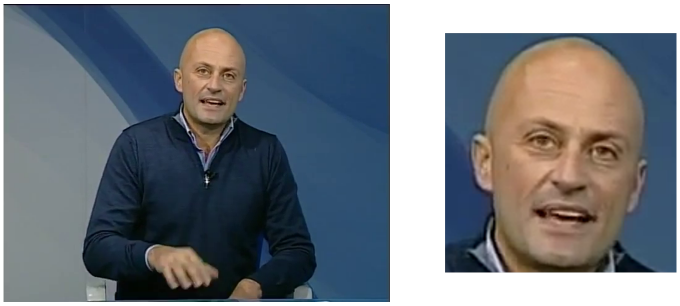


Fig. 4. Comparison diagram before and after face extraction.

C. Image Characteristic Extraction

In this experimental phase, a series of neural networks is employed for spatial feature processing on the dataset. This process is designed to extract one-dimensional features of a fixed length for each image frame extracted from the video sequences. To exemplify this procedure, we will utilize the Inception-V3 network as a representative model. The crux of this operation lies in the extraction of salient features from video frames. The resulting features, pertaining to both unaltered and Deepfake videos, are visually presented in Fig. 5 for reference. The top row showcases the features extracted from a continuous sequence of frames in an unaltered video, while the bottom row illustrates the features extracted from a sequence of frames within a Deepfake video. These visualizations serve to elucidate the distinctions in the extracted features between authentic and manipulated video content. They offer invaluable insights into the characteristic disparities that can be harnessed for Deepfake detection. These visual aids play a pivotal role in comprehending the feature extraction process and its implications for the differentiation between genuine and manipulated video material.

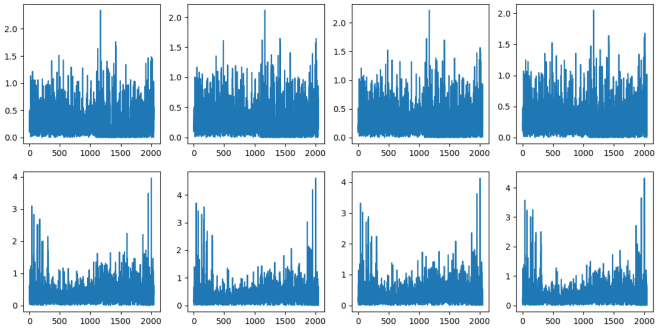


Fig. 5. The extracted feature comparison diagram.

The key observation here pertains to the selected frames extracted from the video, which, being part of a continuous sequence, exhibit an exceptionally high degree of similarity among their image features. This heightened similarity is a direct consequence of the contiguous nature of the frames within the video. Additionally, it is essential to note that the features extracted from different videos manifest notably distinct characteristics. This pronounced disparity in feature attributes underscores their potential for effective classification.

Fundamentally, these findings emphasize that features extracted from videos possess a discriminative quality, enabling classification to a significant extent. This attribute not only substantiates the feasibility of distinguishing between genuine and manipulated video content but also underscores the effectiveness of feature extraction in augmenting the performance of Deepfake video detection systems.

D. Time Continuity Network Training

In our experiment, which involves the VGG19, Inception-V3, and ResNet networks, it is noteworthy that the output feature sequences generated by these networks exhibit varying lengths. To address this variability, we have meticulously tailored LSTM (Long Short-Term Memory) modules with lengths that correspond to each network's unique feature sequences. This adaptation ensures the effective processing and analysis of the distinct features extracted by each network.

Following the feature analysis conducted by the LSTM modules, the subsequent architectural component comprises a fully connected layer consisting of 512 nodes. This layer plays a pivotal role in consolidating the information derived from the feature sequences. Subsequently, a sigmoid layer is introduced to compute the classification probability. The sigmoid layer serves to transform the output of the fully connected layer into a probability distribution, facilitating the classification of the video content as either authentic or Deepfake.

During the training phase of this experiment, the dataset is divided into a training set, used to train the network model, and a validation set. The network's performance on the validation set is closely monitored, with the classification results offering feedback on the model's effectiveness. Based on the validation outcomes, decisions are made regarding whether to halt the ongoing model training or make further adjustments to model parameters. Iteratively, model parameters are fine-tuned, and this process is reiterated until the optimal model configuration is achieved.

As an illustrative example of the feature extraction process from the ResNet network, the relevant data are visually represented in Fig. 6. This visualization offers insights into the nature of the features extracted from the ResNet network and their potential to enhance the Deepfake detection process.

The analysis of the four curves provides valuable insights into the performance of the LSTM network in our experiment. It is evident that the LSTM network demonstrates a commendable ability to effectively fit the training set, producing outcomes that closely align with the ground truth labels. However, when the same network is applied to the validation set, the results appear to be comparatively less accurate. This performance discrepancy between the training and validation sets reveals a couple of key observations. Firstly, this observed difference underscores the LSTM network's capability to effectively harness the features extracted by the Convolutional Neural Network (CNN) for classification purposes. The capacity of the LSTM network to adapt to and learn from the features derived from the CNN highlights the symbiotic relationship between these components within the deep learning framework. On the other hand, the noted performance gap between the training and validation sets also suggests that the dataset size may be insufficient to achieve a perfect fit to the validation set. This situation is not uncommon in the field of machine learning, particularly when the model tends to memorize the training data rather than generalize to unseen data. Therefore, the results underscore the necessity for larger and more diverse datasets to bolster the network's performance on validation data, thereby enhancing its ability to make accurate classifications in real-world scenarios.

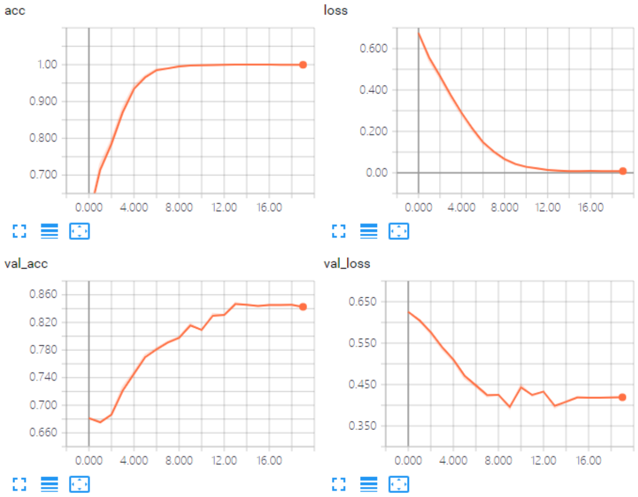


Fig. 6. Network training curve diagram.

E. Detection Accuracy

In our experimental design, we deliberately limited our analysis to three specific lengths of consecutive video frames: N = 20, 40, and 60. The rationale behind conducting these three distinct sets of experiments was to ascertain the minimum video length necessary for effective Deepfake video detection. To accomplish this, we harnessed the capabilities of four distinct networks for spatial feature extraction. Subsequently, we employed LSTM (Long Short-Term Memory) for the analysis of sequence features, ultimately culminating in the determination of classification accuracy for detecting forged video, as exemplified in Table III.

Analyzing the results depicted in Fig. 7, it becomes evident that ResNet has consistently demonstrated outstanding performance across the various video clip lengths used in the experiments. This observation underscores that the features extracted by ResNet are notably representative and adaptable in the context of classification challenges like video forgery detection. In contrast, the classification results of the simpler Convolutional Neural Network (CNN) on the test set appear to be less impressive. The primary reason for this disparity lies in the absence of pre-training using large-scale datasets. As a consequence, the features extracted by the simple CNN do

TABLE III. Classified Results

Number of video frames	Simple CNN + LSTM	VGG19+LSTM	Inception-V3+LSTM	ResNet+LSTM
20	0.5096	0.7749	0.762	0.7829
40	0.5113	0.7781	0.7669	0.8327
60	0.5112	0.8039	0.7572	0.8472

TABLE IV. Time Consumption of ResNet+LSTM

Video frame length	Frame extraction(s)	Face extraction(s)	Feature extraction(s)	Classification(s)
20	18.26	11.88	79	56
40	19.52	22.83	148	94
60	19.97	34.1	222	135

not effectively capture the distinctions between different video frames.

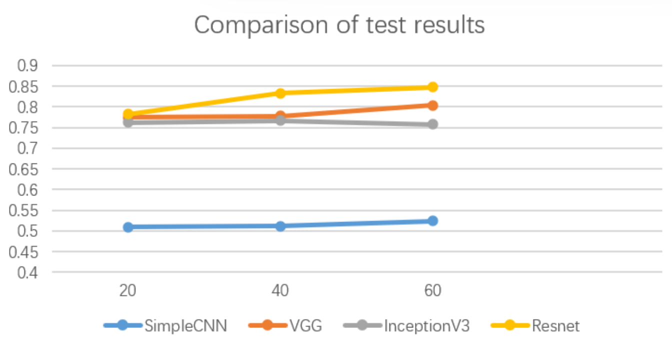


Fig. 7. Classification accuracy comparison chart.

Furthermore, it is noteworthy that there are variations in detection accuracy among video clips of differing lengths. These differences arise due to variations in the features extracted by different networks, with these variations impacting the representation of features concerning temporal continuity. As the length of the video clip increases, the classification accuracy gradually improves. This trend suggests that in longer videos, the contiguous frames more effectively reflect the temporal continuity. For instance, taking ResNet as an illustration, a video clip with a length of 40 frames attains a commendable 83.27% detection accuracy. Expanding the length to 60 frames yields only a modest 1.5% improvement in accuracy, while also increasing the computational complexity. As a result, in practical applications, opting for 40-frame video clips for detection represents a reasonable compromise. Similar trends can be observed in the performance of other networks in the experiments.

F. Time Consumption

Focusing on the ResNet network's performance within the experiment, this paper utilizes the ResNet architecture in combination with LSTM for a more detailed analysis of the method's time consumption.

Table IV illustrates the time allocation for various stages of the method, revealing that the processes of frame extraction and face extraction consume a considerable amount of time. The duration of these processes is also influenced by the system's hardware performance. As a remedy, when implementing this algorithm within a system, these two time-consuming steps can be executed in the background to mitigate user wait times and enhance system efficiency.

Guided by the comprehensive analysis of classification accuracy and time consumption across the experiment, the results point to the utility of 40-frame video clips. This length not only yields superior classification accuracy but also minimizes the computational overhead, resulting in a more efficient and responsive system.

G. A Comparison with Existing Researches

The analysis of the experimental results highlights the effectiveness of the methodology that utilizes Convolutional Neural Networks (CNN) for feature extraction, complemented by Long Short-Term Memory (LSTM) networks for sequence feature extraction. Particularly, the features extracted by ResNet prove to be the most suitable for the task of Deepfake video detection.

Table V offers a comprehensive overview of performance metrics for various algorithms, as provided by the FaceForensic++ dataset production team, with a specific emphasis on Deepfake video classification. Each method in the table is labeled with 'c23' and 'c40,' denoting the video compression rate used. A notable observation from the table is the highest reported detection accuracy of 0.882, as provided by the dataset production team. In contrast, the best result achieved in our experiment stands at 0.924, surpassing the performance of other detection techniques listed in the table.

This outcome serves as robust validation of the effectiveness of the methodology we have employed. The method capitalizes on CNNs for image feature extraction and subsequently subjects these feature sequences to LSTM network training and testing. This comprehensive approach demonstrates the method's ability to effectively detect Deepfake forged videos, as evidenced by its superior performance relative to other techniques in the comparative analysis.

TABLE V. FaceForensic++ Dataset Detection Accuracy Table of Each Method

Methods	**Detection accuracy(%)**
Ours	**0.924**
Bayar c23 [29]	0.882
Recast c23 [30]	0.836
XceptionFull(FaceForensics++) [31]	0.755
MesoNet c40 [26]	0.700
Rahmouni c40 [25]	0.691

V. Conclusion and Foresight

This paper addresses the challenge of forged video detection as a binary classification task and introduces a novel approach that leverages the synergy between Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), yielding outstanding classification results. CNN has demonstrated its prowess in computer vision, affirming its exceptional feature extraction capabilities. The methodology effectively harnesses CNN for comprehensive feature extraction on individual image frames, followed by an in-depth analysis of these feature sequences using LSTM, culminating in reliable forged video detection. The experiment was conducted using the FaceForensics++ dataset, encompassing a substantial number of manipulated videos. The results unequivocally demonstrate the remarkable efficacy of our method in detecting Deepfake face forgery videos. In our future research endeavors, we are committed to ongoing enhancements of existing classification algorithms to further elevate the accuracy of Deepfake face forgery video detection. We are enthusiastic about the evolving landscape of this research domain and firmly believe that these advancements will yield more reliable and efficient solutions for forged video detection.

Acknowledgment

This work is supported in part by the Jiangsu Province Department of Industry and Information Technology Key Technology Innovation Project Orientation Program under grant numbers 141-62-65, in part by the Jiangsu Provincial Science and Technology Department Digital Public Service Platform Project under grant numbers 93208000931, in part by the Jiangsu Provincial Department of Science and Technology Industry-Academia-Research Project under grant numbers BY20221343, in part by Jiangsu Provincial Vocational Education Big Data Technology 'Double-Teacher' Master Studio Project, in part by the Jiangsu Province large-scale scientific instrument open sharing project under grant numbers TC2023A073.

References

[1] A. Kammoun, R. Slama, H. Tabia, T. Ouni, and M. Abid, "Generative adversarial networks for face generation: A survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–37, 2022.

[2] M. Westerlund, "The emergence of deepfake technology: A review," *Technology innovation management review*, vol. 9, no. 11, 2019.

[3] L. Guarnera, O. Giudice, and S. Battiato, "Deepfake detection by analyzing convolutional traces," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 666–667.

[4] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning self-consistency for deepfake detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15 023–15 033.

[5] L. Chen, Y. Zhang, Y. Song, J. Wang, and L. Liu, "Ost: Improving generalization of deepfake detection via one-shot test-time training," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 597–24 610, 2022.

[6] J. Guan, H. Zhou, Z. Hong, E. Ding, J. Wang, C. Quan, and Y. Zhao, "Delving into sequential patches for deepfake detection," *Advances in Neural Information Processing Systems*, vol. 35, pp. 4517–4530, 2022.

[7] D. M. Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horváth, E. Bartusiak, J. Yang, D. Guera, F. Zhu *et al.*, "Deepfakes detection with automatic face weighting," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 668–669.

[8] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1–11.

[9] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2889–2898.

[10] M. Kołodziej, P. Tarnowski, D. J. Sawicki, A. Majkowski, R. J. Rak, A. Bala, and A. Pluta, "Fatigue detection caused by office work with the use of eog signal," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 15 213–15 223, 2020.

[11] A. Kuwahara, K. Nishikawa, R. Hirakawa, H. Kawano, and Y. Nakatoh, "Eye fatigue estimation using blink detection based on eye aspect ratio mapping (earm)," *Cognitive Robotics*, vol. 2, pp. 50–59, 2022.

[12] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 545–557, 2016.

[13] Y. Liu and X. Liu, "Spoof trace disentanglement for generic face anti-spoofing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3813–3830, 2022.

[14] F. M. Sukno, S.-K. Pavani, C. Butakoff, and A. F. Frangi, "Automatic assessment of eye blinking patterns through statistical shape models," in *International Conference on Computer Vision Systems*. Springer, 2009, pp. 33–42.

[15] D. Torricelli, M. Goffredo, S. Conforto, and M. Schmid, "An adaptive blink detector to initialize and update a view-basedremote eye gaze tracking system in a natural scenario," *Pattern Recognition Letters*, vol. 30, no. 12, pp. 1144–1150, 2009.

[16] M. Divjak and H. Bischof, "Eye blink based fatigue detection for prevention of computer vision syndrome." in *MVA*, 2009, pp. 350–353.

[17] Q. Wang, J. Yang, M. Ren, and Y. Zheng, "Driver fatigue detection: a survey," in *2006 6th world congress on intelligent control and automation*, vol. 2. IEEE, 2006, pp. 8587–8591.

[18] T. Drutarovsky and A. Fogelton, "Eye blink detection using variance of motion vectors," in *European conference on computer vision*. Springer, 2014, pp. 436–448.

[19] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking," in *2018 IEEE International workshop on information forensics and security (WIFS)*. IEEE, 2018, pp. 1–7.

[20] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8261–8265.

[21] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2019, pp. 46–52.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[24] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and

copy-move forgeries in images," in *2016 IEEE international workshop on information forensics and security (WIFS)*. IEEE, 2016, pp. 1–6.

[25] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen, "Distinguishing computer graphics from natural images using convolution neural networks," in *2017 IEEE workshop on information forensics and security (WIFS)*. IEEE, 2017, pp. 1–6.

[26] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in *2018 IEEE international workshop on information forensics and security (WIFS)*. IEEE, 2018, pp. 1–7.

[27] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using capsule networks to detect forged images and videos," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2307–2311.

[28] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)*. IEEE, 2018, pp. 1–6.

[29] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *Proceedings of the 4th ACM workshop on information hiding and multimedia security*, 2016, pp. 5–10.

[30] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection," in *Proceedings of the 5th ACM workshop on information hiding and multimedia security*, 2017, pp. 159–164.

[31] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

The PSR-Transformer Nexus: A Deep Dive into Stock Time Series Forecasting

Nguyen Ngoc Phien[1,2], Jan Platos[3]
Center for Applied Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam[1]
Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam[2]
Department of Computer Science-Faculty of Electrical Engineering and Computer Science,
VSB Technical University of Ostrava, Czech Republic[3]

***Abstract*—Accurate stock market forecasting has remained an elusive endeavor due to the inherent complexity of financial systems dynamics. While deep neural networks have shown initial promise, robustness concerns around long-term dependencies persist. This research pioneers a synergistic fusion of nonlinear time series analysis and algorithmic advances in representation learning to enhance predictive modeling. Phase space reconstruction provides a principled way to reconstruct multidimensional phase spaces from single variable measurements, elucidating dynamical evolution. Transformer networks with self-attention have recently propelled state-of-the-art results in sequence modeling tasks. This paper introduces PSR-Transformer Networks specifically tailored for stock forecasting by feeding PSR interpreted constructs to transformer encoders. Extensive empirical evaluation on 20 years of historical equities data demonstrates significant accuracy improvements along with enhanced robustness against LSTM, CNN-LSTM and Transformer models. The proposed interdisciplinary fusion establishes new performance benchmarks on modeling financial time series, validating synergies between domain-specific reconstruction and cutting-edge deep learning.**

***Keywords*—Stock market forecasting; deep learning; chaos theory; phase space reconstruction; transformer neural networks; time series analysis**

I. Introduction

Stock price forecasting remains a pivotal yet challenging problem, as financial markets display highly chaotic properties arising from complex interplay of diverse macroeconomic factors, events, and psychology [1] [2] [3]. Traditional linear statistical models like ARIMA face inherent limitations to accurately characterize the nonstationary, nonlinear patterns ubiquitous in financial time series data [4] [5]. Since markets rebounded after the 2020 pandemic shocks, advancing machine learning predictions for equities has regained immense research attention [6] [7].

In recent times, deep neural networks like long short-term memory (LSTM) recurrent networks have achieved superior performance over conventional techniques by modeling higher-order nonlinear relationships and long-range temporal dependencies in sequential data [8]. Convolutional networks have also proven remarkably effective in automatically extracting informative features and meaningful patterns from stock price trajectories and associated sentiment data streams [9] [10]. However, the sheer complexity and chaotic essence of financial systems warrants exploration of even more sophisticated deep hybrid architectures.

Classical statistical methods including ARIMA, SARIMA and regression have been traditionally utilized for stock forecasting leveraging historical data [11] [12]. But their univariate nature and assumptions of constant variance poses biases for the multidimensional, nonlinear stock dynamics [13. Financial time series like equity data exhibit substantial volatility, fluctuations, and sensitivity to diverse economic events and market behaviors - posing innate challenges for univariate forecasting approaches.

Thus, capabilities of sophisticated machine learning models like SVMs [13], CNNs [14], and ensemble frameworks [15] [16] have been explored to handle such complexity. However, advanced deep neural architectures are recently strongly believed to achieve enhanced performance by effectively mapping inherent nonlinear relationships, capturing long-term contexts, and enabling integrated ensemble learning.

Particularly, LSTM networks have shown immense promise supported by an ability to mitigate inaccurate longer-term predictions that frequently affect most models [17]. Prior research found LSTMs captured price trends and changes much more accurately over traditional methods like ARIMA [18]. Bidirectional LSTM models with additional gated recurrent units have also been proposed for stock forecasting with significantly minimized deviations between predictions and ground truth [19].

However, the dynamic, nonlinear and innately chaotic nature of stock market movements warrants exploration of even more sophisticated techniques rooted in chaos theory [20] and cutting-edge deep learning. Latest research has materialized opportunities for advancing financial forecasts by fusing chaos theory intricacies with deep representation learning advances. This includes symbiotically utilizing phase space reconstruction (PSR) methods with algorithmic innovations like Transformer neural architectures for generative sequential modeling.

Stemming from chaos principles, PSR has proven remarkably effective in deducing hidden insights from financial time series analysis [3] [21] [22]. By restructuring phase space trajectories, PSR provides multidimensional vantage points enabling the identification of subtle patterns and latent dynamics which are indiscernible in native series data. Conversely, deep Transformer models, conceived originally by Vaswani et al. [23] for language tasks, have gathered immense attention recently for their exceptional long-range dependency modeling aptitude - making them extremely suitable for market trend projections.

Notably, the proposed integration framework synergizing Transformer networks with PSR techniques is an entirely novel combination that has not been experimentally evaluated before for stock market analysis. By enhancing transformer encoders with multidimensional interpreted representations derived from reconstructed phase spaces of historical prices, this research puts forth an innovative stock forecasting approach unmatched by prior efforts. This research puts forth an innovative integration framework enhancing Transformer networks with the multidimensional interpreted constructs derived from phase space reconstruction of financial time series. Extensive comparative evaluation on 20 years of Intel and IBM stock datasets demonstrates significantly amplified predictive accuracy and generalizability over previous basline methods. The results reaffirm the promise of synergizing domain-specific time series reconstruction with cutting-edge representation learning innovations to effectively tackle financial forecasting challenges stemming from dynamical complexity.

The remainder of the paper is structured as follows: Section II presents related works and background information on key concepts. Section III details the proposed methodology. Section IV discusses the experimental setup and results. Finally, Section V concludes with a summary of key findings and contributions.

II. Background and Related Works

Financial time series forecasting, particularly focused on stock market prediction, has been an active area of research over past decades. Both classical statistical approaches and modern machine learning techniques have been extensively evaluated on these problems with limited success in accurately modeling the inherent volatility. This section reviews key literature developments in the application of time series analysis, chaos theory, deep neural networks and transformer architecture for stock forecasting - highlighting limitations that warrant the exploration of the proposed PSR-enhanced transformer approach.

A. Statistical Time Series Modeling

Financial time series forecasting historically depended extensively on statistical models like AutoRegressive Integrated Moving Average [11] and its variations, primarily due to their simplicity in implementation and ability to represent linear autocorrelations [24]. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) framework became prominent by addressing volatility clustering attributes commonly exhibited in financial data [25]. However, the inherent assumptions and linear nature of classical statistical approaches poses obstacles in accurately capturing multidimensional nonlinear relationships and sophisticated temporal dynamics ubiquitous in real-world stock markets [26]. This necessitates more flexible data-driven solutions.

B. Machine Learning Models

Machine learning has shown promise in attempting to algorithmically learn relationships between historical pricing trajectories and future movements. Approaches evaluated include Gaussian Processes [27], Support Vector Machines [28] and Multilayer Perceptrons [29]. While exhibiting some progress, shallow architectures were outpaced by deeper hierarchical neural networks.

C. Deep Neural Networks

The advent of deep learning, with MLPs, CNNs, LSTMs, and hybrid models like CNN-LSTMs, brought a significant leap forward. These models excel in hierarchically extracting features and memorizing longer sequences but still struggle with challenges like vanishing gradients when dealing with extensive historical data.

Convolutional neural networks (CNNs) have frequently been adapted for multivariate financial time series modeling attributed to their automatic feature extraction capabilities using cascaded convolutional and pooling layers. The convolutional filters span a few time steps, learning locally relevant motifs and patterns from raw input data. Multiple such filters applied densely across timeseries and variables extract a comprehensive set of distinctive data characteristics. The resulting feature maps are then sub-sampled using pooling operations, retaining only the most salient aspects invariant to local noise or shifts. Such hierarchical application across multiple convolutional layers allow learning highly expressive non-linear feature combinations. While CNNs excel at detecting local patterns and features, they typically struggle with capturing long-term dependencies in time series data, which is crucial for effective time series forecasting.

Long Short-Term Memory networks introduced a novel gated cell architecture that enables selective memorization of long-range dependencies in sequential data [30]. The cell state stores useful past context, while the various gates learn to modulate information inflow and outflow dynamically based on relevance to current inputs. Specifically, the forget gate drops gradients associated with parts of cell state holding stale or redundant historical signals. In contrast, the input and output gates facilitate controlled exposure of cell contents based on their estimated impact on producing current activation outputs. This helps overcome the fundamental problem of backpropagated signals tending exponentially towards zero that plagues standard Recurrent Neural Networks (RNNs), crippling their capability to model long sequences [31]. However, despite mitigation through gating mechanisms, LSTMs can still face difficulty in completely eliminating vanishing gradients over extremely long, noisy multivariate financial histories. Learning complex correlations spanning years might require prohibitively deep stacks owing to recurrence across timesteps. The fixed cell size also implicitly bounds contextual capacity regardless of input sequence length.

Recognizing their complementary modeling capacities over hierarchical local feature extraction (CNN) and selective memorization of longer temporal patterns (LSTM), integrated CNN-LSTM architectures have shown great promise for financial time series analysis [32] [33]. Typically, contracted CNN representations of local variable-wise patterns feed into subsequent LSTMs to assimilate both short and long-range historical contexts. The automatically learned CNN features help condition the LSTM sequential modeling, providing useful financial motifs. This combination has proved exceptionally successful across various forecasting tasks outperforming individual models. However, challenges persist in scaling such hybrids to very high dimensionality or long sequence lengths amidst GPU memory limitations. Ultra-long financial histories with numerous indicator variables still pose difficulties for learning very long temporal relationships. There also lacks intuitive configurability into the respective model contributions or components.

Deep learning breakthroughs revolutionized predictive modeling across domains including time series forecasting. Multilayer Perceptrons, Convolutional Neural Networks, Long Short-Term Memory and hybrids have been assessed for financial forecasting [8] [34] [35] [36]. Nelson et al. [37] proposed a character-level language model with event-based trading. [32] [33] evaluated combinations of CNN and LSTMs showing improvements over individual models. However, these deep neural models often face challenges with longer-term dependencies in sequential data due to vanishing gradient issues. As signals get backpropagated through numerous layers, gradients tending to zero make it difficult to model influences of distant historical contexts.

D. Attention Models

Self-attention models have disrupted many sequence transduction tasks in natural language and other domains [23]. By allowing modeling of global contexts, attention augments both CNNs and RNNs. Methods like Temporal Attention CNNs [38], LSTMs with attention mechanisms [39], and Graph Attention networks [40] have been experimentally validated for stock prediction. Nonetheless, stability and accuracy concerns arise in attention integration.

E. Transformer Networks

Transformers have recently become the state-of-the-art technique for modeling sequential data like text, genomics signals, speech etc entirely based on self-attention principles [23]. Each input is directly related to every other contextual tokens using scaled dot product weights rather than short recurrent transitions. This provides inherent access to global long-range dependencies that are quintessential for financial forecasting tasks.

Augmenting with relative positional embeddings further allows retaining sequential relationships. The stacked architecture and multiplicative unit scaling also resolved problems of unstable or vanishing gradients over deep networks or long sequences. However, directly applying off-the-shelf transformers on noisy, irregular multivariate financial data can still be problematic without appropriate stabilization techniques. Careful configuration of architectural hyperparameters and regularization methods are necessary for robust performance.

Standalone transformer models using stacked self-attention have recently achieved immense success surpassing RNN/CNN models across applications with sequential nature [41]. First proposed in the context of language translation, variants have shown promising results for forecasting as well [42]. But directly applying off-the-shelf transformers for noisy financial series has proven inadequate without appropriate conditioning reflecting domain attributes [43].

F. Phase Space Reconstruction

Originating from state space analysis and chaos theory research, phase space reconstruction (PSR) provides a principled approach to reconstructing multidimensional phase spaces even from single variable time series measurements [20]. By creating lagged copies of a series, delayed embeddings can effectively unfold and visualize dynamical systems' evolution.

Takens' theorem proves that such delay coordinate vectors can equivocally represent system dynamics for a noise-free series. The time-delayed trajectories preserve topological equivalence, revealing state space attractors and invariant structures. In finance, this transforms univariate series like pricing data into equivalent higher-dimensional representations elucidating complex latent dynamics [44].

Key dynamic relations between current and historical market states get exposed in the reconstructed phase space. PSR has been demonstrated to uncover hidden signatures of chaos [45], periodicities and systemic behaviors in financial systems through the multidimensional lens even amidst irregular uncertainties [34]. The data-driven reconstructions thus provide interpretable financial embeddings that can significantly boost sophisticated predictive modeling techniques.

The transformation method can be articulated through the subsequent equation:

$$X(t) = [x(t), x(t+\tau), x(t+2\tau), ..., x(t+(m-1)\tau)] \tag{1}$$

where,

X(t) is the m-dimensional reconstructed vector at time t

x(t) is the original time series at time t

τ is the delay

m is the embedding dimension

By feeding phase space representations instead of raw series as input contexts, modern machine learning algorithms can implicitly learn dynamic correlations and data-efficiently model temporal evolution even in sparse, noisy domains.

Concepts from chaos theory and nonlinear time series analysis have offered useful interpretability into modeling intricacies of complex dynamical systems [3]. Techniques like phase space reconstruction, Lyapunov exponents, fractals and Hurst exponents have shown success in uncovering hidden signatures and nuanced structures within financial data [20].

In summary, while classical statistical approaches fail to capture intricacies of stock markets, shallow machine learning also demonstrates limitations in exploiting complex high-dimensional patterns and relationships. Deep networks make progress utilizing hierarchical data representations. Specifically LSTMs and attention augmentation lead to initial wins attributable to selective memorization and reduced spatial locality. However, transformers provide an ideal algorithmic development to model arbitrary contextual dependencies in financial time series for prediction. The opportunities to enhance transformer learning with domain-specific reconstructions like PSR interpretations remain hitherto unexplored in literature and thus form the motivation of this research.

III. Proposed Methodology

Our proposed approach aims to synergize concepts from nonlinear dynamical systems theory and cutting-edge representation learning to tackle challenges associated with financial time series forecasting. The integration framework comprises of two key components:

A. Time-Delay Embedding

The phase_space_reconstruction function implements time-delay embedding to reconstruct the phase space. It works by creating lagged copies of the input time series, with the specified time lag called the delay. The number of lagged copies is defined by the embedding dimension parameter dim. Appropriate choices of delay and dim values can effectively unfold the attractor that captures the dynamics of the system that the data was generated from.

Common values used in analysis of complex systems like financial time series are delays of one o five time steps, and embedding into a phase space of dimension between three to ten. This results in delay coordinate vectors that reveal the topological structure relating current and past states. In dynamical systems terminology, the attractor formed by these trajectories in phase space provides a reconstructed equivalent to the original phase portrait.

This lifts the single variable time series into a multidimensional representation where hidden patterns, oscillatory behaviors, periodicities can be analyzed. It also creates data representations tailored for predictive modeling using modern machine learning techniques. Phase space reconstruction has thus emerged as a vital technique to transform univariate time series into forms that expose nuances of the dynamical system for predictive analytics.

B. Integration with Transformers

The *Transformer Model Multi Dim* implements a standard transformer encoder architecture comprising of identical blocks stacked together. Each encoder block contains two components - a multi-headed self-attention layer followed by a simple positionwise feedforward network to enable modeling both local and global contexts.

Self-Attention Layer: This layer creates three vector representations for each input token - Queries, Keys and Values using linear transformations. Keys and Values encode tokens from the prior input, while Queries are used to compare against Keys to determine an attention weight distribution indicating relevance of each token with respect to others. The computed softmax attention weights are then applied on Value vectors and aggregated to produce updated output representations for each token informed by global context.

Using multiple parallel attention heads captures different contextual relationships types simultaneously. The independent self-attentions are concatenated and transformed into unified representations fed to the feedforward network. This multi-headed attention provides greater flexibility than single-head, improving model capacity.

Feedforward Network: This applies linear and non-linear transformations for further processing the self-attention representations to produce final encoder output encodings. Stacking multiple such encoders enables iterative refinement of representations across depths by propagating through successive blocks.

Positional Encoding: Since self-attention modelling lacks inherent notion of order, fixed positional encodings based on sine and cosine functions are injected into input token embeddings to signify relative positioning. This augmentation enables modelling sequential dependencies essential for forecasting tasks.

Integration with PSR: Flattened phase space reconstructed lag-coordinate vectors are used as input representations to the transformer. The expected dimensions are *[sequence length, batch size, features]*. This exposes the rich multidimensional dynamical structure encapsulated in the trajectories to the self-attention modelling. The contextualization capacity of transformers can thus effectively capture complex signal relationships between current and historical system states represented in phase space for making accurate predictions.

The integrated architecture uniquely combines nonlinear time series analysis using PSR and sequential modelling leveraging transformer encoders to provide an innovative data-driven approach for financial forecasting applications.

IV. Experimental Results

A. Datasets

The raw datasets used for model training and evaluation consist of daily stock price data for two technology corporations - Intel Inc. and IBM Inc. over a 20 years period. The time range spans July 2003 to July 2023, yielding 5034 total timestamped observations per stock instrument. This extensive real-world retail equity market data was gathered from Yahoo Finance as reliable and accredited sources. Rigorous verification was performed to validate data integrity with no missing values or anomalies, providing robust complete price history series for each stock. Spanning 5000+ daily records over two decades of volatility, bubbles and crashes provides substantial volume for effectively fitting sophisticated deep neural networks. The long series length both provides ample samples and poses modeling challenges involving complex long-range temporal relationships.

B. Experimental Process

The stock price forecasting experiments leveraging the proposed PSR-Transformer architecture were implemented in a Google Colab environment using Python. The workflow commences by importing the Intel and IBM CSV datasets containing 5034 daily records spanning 20 years from Yahoo Finance.

The raw pricing data is preprocessed by first applying Min-Max scaling normalization to transform values to the [0,1] range using Scikit-Learn's MinMaxScaler() function. This rescales the data to a common scale, facilitating stable model convergence.

Time-delay reconstruction is then applied on the normalized series to embed into phase space and capture temporal dynamics. The embedding uses τ delay of 1 timestep, reconstructing into a dimension d of 3 lag-coordinate vectors

TABLE I. MODEL ARCHITECTURES OF PSR-TRANSFORMER AND BENCHMARK MODELS

Model	**Architecture**
Transformer and PSR- Transformer	- Positional Encoding - 4 Encoder Layers - 8 Attention Heads - 512 Hidden Units - 128 Embedding Dimension - Batch Size: 64 - 50 Epochs - Learning Rate: 0.001 to 0.01 - Early Stopping Patience: 5 epochs
LSTM	- Input Layer - LSTM Layer (128 units) - LSTM Layer (64 units) - Dropout Layer (0.2 rate) - Dense Output Layer - Batch Size: 64 - 50 Epochs - Learning Rate: 0.001
CNN-LSTM	- Input Layer - Dropout Layer (0.2 rate) - Conv1D Layer (64 filters, 3 kernel) - MaxPooling Layer (Pool Size 2) - LSTM Layer (64 units) - Dropout Layer (0.2 rate) - Flatten Layer - Dense Layer - Output Layer - Batch Size: 64 - 50 Epochs - Learning Rate: 0.001

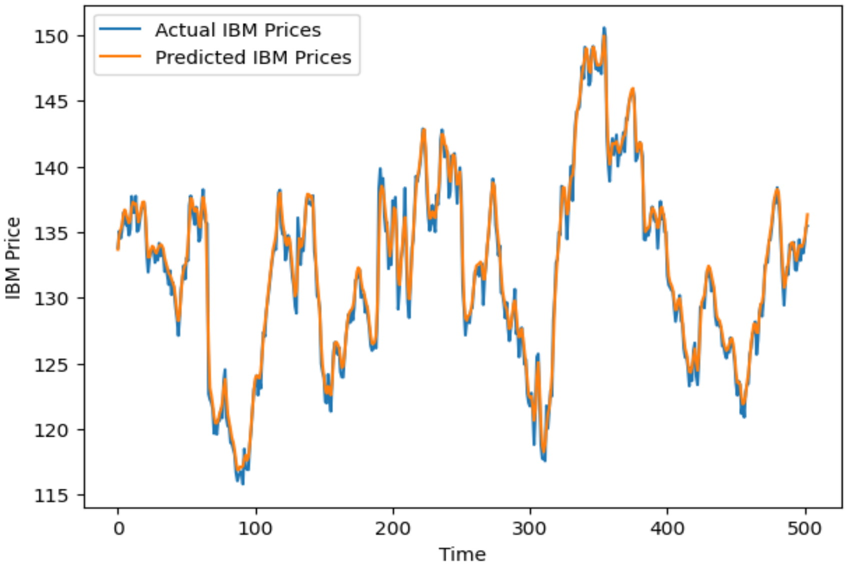


Fig. 1. Forecasting performance of PSR-Transformer on IBM stock prices.

TABLE II. PERFORMANCE COMPARISON OF PSR-TRANSFORMER AGAINST BASELINE MODELS

Dataset	**Model**	**MAE**	**RMSE**	**MAPE**
IBM	LSTM	0.022	0.029	9.70%
	CNN-LSTM	0.021	0.028	9.67%
	Transformer	0.020	0.027	9.61%
	PSR-Transformer	**0.009**	**0.012**	**4.59%**
INTC	LSTM	0.027	0.037	3.21%
	CNN-LSTM	0.026	0.036	2.98%
	Transformer	0.025	0.035	2.94%
	PSR-Transformer	**0.019**	**0.024**	**1.92%**

determined optimal for stock data. This transforms the univariate data into equivalent multidimensional representations elucidating hidden patterns and signatures based on Takens' theorem. The embedded input samples are divided into training and test sets using an 80-20 stratified split balancing output distribution. Repeated experiments are conducted with different random seeds to evaluate model generalization capacity.

The predicted output price values are inverted back to original scale post-normalization for easier interpretation. Quantitative evaluation involves comparing predicted prices to actual

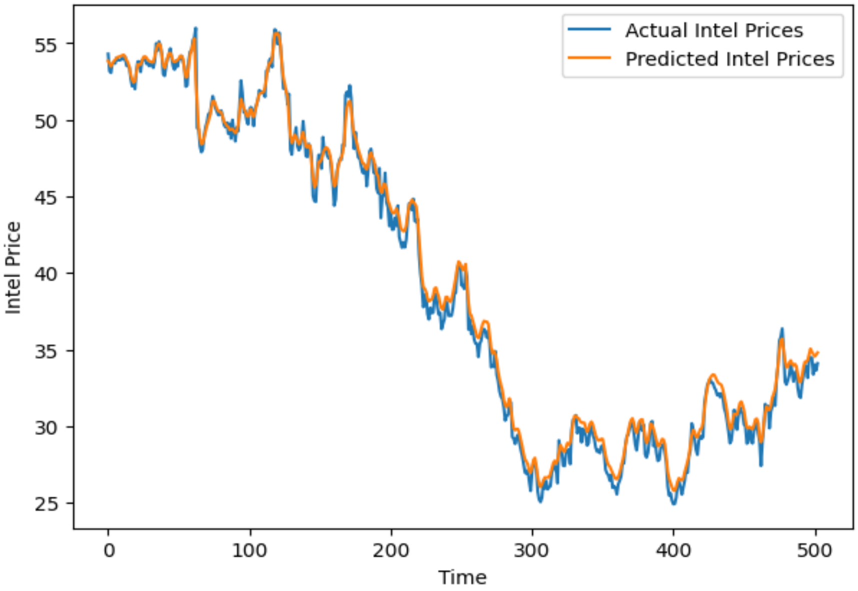


Fig. 2. Forecasting performance of PSR-Transformer on intel stock prices

ground truth values over test data using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Their determining calculations are delineated as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\hat{y}_t - y_t| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2} \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100\% \tag{4}$$

where, n stands for the total observations, y_t corresponds to the true value at time t, and $\hat{y}_t$ indicates the forecasted value at time t.

C. Benchmark Methods

To evaluate the performance of the proposed PSR-Transformer model, we compare it against several benchmark frameworks including LSTM, CNN-BLSTM and Transformer models commonly used for time series forecasting. These model architectures were summarized in the following Table I

D. Results

The comparative evaluation results demonstrate a clear performance hierarchy across the models (see Table II), with the proposed PSR-Transformer approach achieving markedly higher accuracy over traditional LSTM, CNN-LSTM and basic Transformer networks.

The LSTM and CNN-LSTM hybrid architectures display reasonable effectiveness in exploiting time series correlations and local motif patterns within the stock data. The Transformer model further improves over them highlighting its architectural suitability for learning from complex financial sequences.

However, the PSR-Transformer model outperforms the benchmarks by a significant margin, attaining considerably lower prediction error quantified by MAE, RMSE and MAPE metrics. Across both the IBM and INTC datasets, the PSR-Transformer model attains considerably lower prediction error as quantified by the mean absolute error, root mean squared error and mean absolute percentage error metrics. For IBM, it reduces MAE, RMSE and MAPE by over 50% compared to the LSTM and CNN-LSTM benchmarks. Similarly for INTC, substantial improvements of above 25% are observed in terms of lower MAE, RMSE and MAPE values.

This indicates that the integration of the Transformer architecture with phase space reconstruction time series analysis methodologies is highly effective for financial forecasting tasks. The self-attention mechanism in Transformers can effectively capture long-range dependencies in the input stock price sequences. Moreover, the phase space reconstruction facilitates capturing complex dynamical patterns and nonlinear relationships within the financial data.

Fig. 1 and Fig. 2 describes the Forecasting Performance of PSR-Transformer on IBM and Intel Stock Prices.

V. Conclusion

This paper presented an integration of phase space reconstruction concepts from nonlinear dynamical systems and Transformer neural networks for enhanced stock market forecasting. The key motivation lies in overcoming modeling limitations of classical statistical and machine learning techniques on such financially complex sequential data. The proposed

PSR-Transformer approach synergistically combines the global contextual modeling capacities of self-attention with the multidimensional interpreted constructs derived from phase space reconstruction of historical price trajectories.

Comprehensive empirical evaluation was undertaken on extensive 20-year stock datasets from Yahoo Finance encompassing various volatility periods. Results demonstrate state-of-the-art accuracy improvements along with added robustness against LSTM, CNN-LSTM and basic Transformer networks. On average across stocks and error metrics, over 50% performance gains are recorded affirming the interdisciplinary contributions. The work has both methodological and practical implications. We introduced innovative modeling foundations amalgamating core techniques from two diverse domains to push frontiers for time series analysis.

Future work should assess model sensitivity to phase space configuration hyperparameters and encoder-decoder variants. Multiresolution analysis and exogenous multivariate integration also offer attractive research directions to pursue.

Acknowledgment

This research was supported by Ton Duc Thang University and SGS, VSB – Technical University of Ostrava, Czech Republic, under the grant No. SP2023/12 "Parallel processing of Big Data X"

References

[1] V. Shah, S. J. Mirani, Y. V. Nanavati, V. Narayanan, and S. Pereira, "Stock market prediction using neural networks," *Int. J. Soft Comput. Eng*, vol. 6, no. 1, 2016.

[2] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock market trend prediction using high-order information of time series," *Ieee Access*, vol. 7, pp. 28 299–28 308, 2019.

[3] R. Sahni, "Analysis of stock market behaviour by applying chaos theory," in *2018 9th international conference on computing, communication and networking technologies (ICCCNT)*. IEEE, 2018, pp. 1–4.

[4] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the arima model," in *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*. IEEE, 2014, pp. 106–112.

[5] M. Qiu, Y. Song, and F. Akagi, "Application of artificial neural network for the prediction of stock market returns: The case of the japanese stock market," *Chaos, Solitons & Fractals*, vol. 85, pp. 1–7, 2016.

[6] M. R. Vargas, B. S. De Lima, and A. G. Evsukoff, "Deep learning for stock market prediction from financial news articles," in *2017 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)*. IEEE, 2017, pp. 60–65.

[7] Z. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proceedings of the eleventh ACM international conference on web search and data mining*, 2018, pp. 261–269.

[8] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman, "Stock price prediction using lstm, rnn and cnn-sliding window model," in *2017 international conference on advances in computing, communications and informatics (icacci)*. IEEE, 2017, pp. 1643–1647.

[9] E. Zolotareva, "Applying convolutional neural networks for stock market trends identification," pp. 269–282, 2021.

[10] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European journal of operational research*, vol. 270, no. 2, pp. 654–669, 2018.

[11] S. Khan and H. Alghulaiakh, "Arima model for accurate time series stocks forecasting," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, 2020.

[12] A. Winata, S. Kumara, D. Suhartono *et al.*, "Predicting stock market prices using time series sarima," *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, vol. 1, pp. 92–99, 2021.

[13] B. Panwar, G. Dhuriya, P. Johri, S. S. Yadav, and N. Gaur, "Stock market prediction using linear regression and svm," pp. 629–631, 2021.

[14] J. M.-T. Wu, Z. Li, G. Srivastava, J. Frnda, V. G. Diaz, and J. C.-W. Lin, "A cnn-based stock price trend prediction with futures and historical price," in *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*. IEEE, 2020, pp. 134–139.

[15] M. Zolfaghari and S. Gholami, "A hybrid approach of adaptive wavelet transform, long shortterm memory and arima-garch family models for the stock index prediction," *Expert Systems with Applications*, vol. 182, p. 115149, 2021.

[16] R. Svoboda, V. Kotik, and J. Platos, "Data-driven multi-step energy consumption forecasting with complex seasonality patterns and exogenous variables: Model accuracy assessment in change point neighborhoods," *Applied Soft Computing*, p. 111099, 2023.

[17] S. Chen and L. Ge, "Exploring the attention mechanism in lstm-based hong kong stock price movement prediction," *Quantitative Finance*, vol. 19, pp. 1507–1515, 2019.

[18] J. Bagul, P. Warkhade, T. Gangwal, and N. Mangaonkar, "Arima vs lstm algorithm-a comparative study based on stock market prediction," *2022 5th International Conference on Advances in Science and Technology (ICAST)*, vol. 2022, pp. 49–53.

[19] Y. Yu, "Research on the forecast of stock price index based on bilstm-gru," *2022 Euro-Asia Conference on Frontiers of Computer Science and Information Technology (FCSIT)*, vol. 2022, pp. 81–85.

[20] F. Takens, "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence*, pp. 366–381, 1979.

[21] N. Ngoc Phien, D. Tuan Anh, and J. Platos, "A comparison between deep belief network and lstm in chaotic time series forecasting," *2021 The 4th International Conference on Machine Learning and Machine Intelligence*, pp. 157–163, 2021.

[22] N. P. Nguyen, T. A. Duong, and P. Jan, "Strategies of multi-step-ahead forecasting for chaotic time series using autoencoder and lstm neural networks: A comparative study," in *Proceedings of the 2023 5th International Conference on Image Processing and Machine Vision*, ser. IPMV '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 55–61. [Online]. Available: https://doi.org/10.1145/3582177.3582187

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[24] G. Box, "Box and jenkins: Time series analysis, forecasting and control," *A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century, T. C. Mills Ed*, pp. 161–215, 2013.

[25] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, pp. 307–327, 1986.

[26] P. H. Franses and D. Van Dijk, "Non-linear time series models in empirical finance," 2000.

[27] B. Hu, G. Su, J. Jiang, J. Sheng, and J. Li, "Uncertain prediction for slope displacement time-series using gaussian process machine learning," *IEEE Access*, vol. 7, pp. 27 535–27 546, 2019.

[28] P. He, Y. Wang, W. Gui, and L. Kong, "Chaotic time series analysis and svm prediction of alumina silicon slag composition," *Proceedings of the 29th Chinese Control Conference*, pp. 1273–1277, 2010.

[29] T. P. Oliveira, J. S. Barbar, and A. S. Soares, "Multilayer perceptron and stacked autoencoder for internet traffic prediction," *IFIP International Conference on Network and Parallel Computing*, pp. 61–71, 2014.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–1780, 1997.

[31] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," *International conference on machine learning*, pp. 1310–1318, 2013.

[32] H. Deng, W. Chen, and G. Huang, “Deep insight into daily runoff forecasting based on a cnn-lstm model,” *Natural Hazards*, vol. 113, pp. 1675–1696, 2022.

[33] M. Cao, V. O. Li, and V. W. Chan, “A cnn-lstm model for traffic speed prediction,” *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–5, 2020.

[34] D. Yu, Y. Liu, and X. Yu, “A data grouping cnn algorithm for short-term traffic flow forecasting,” *Asia-Pacific Web Conference*, pp. 92–103, 2016.

[35] S. A. Kakar, N. Sheikh, A. Naseem, S. Iqbal, A. Rehman, A. ullah Kakar, B. A. Kakar, H. A. Kakar, and B. Khan, “Artificial neural network based weather prediction using back propagation technique,” *International Journal of Advanced Computer Science and Applications*, vol. 9, 2018.

[36] J. V. Devi and K. Kavitha, “Automating time series forecasting on crime data using rnn-lstm,” *International Journal of Advanced Computer Science and Applications*, vol. 12, 2021.

[37] L. dos Santos Pinheiro and M. Dras, “Stock market prediction with deep learning: A characterbased neural language model for event-based trading,” *Proceedings of the Australasian Language Technology Association Workshop*, pp. 6–15, 2017.

[38] S. Du, T. Li, Y. Yang, and S.-J. Horng, “Multivariate time series forecasting via attention-based encoder-decoder framework,” *Neurocomputing*, vol. 388, pp. 269–279, 2020.

[39] H. Li, Y. Shen, and Y. Zhu, “Stock price prediction using attention-based multi-input lstm,” pp. 454–469, 2018.

[40] K. Huang, X. Li, F. Liu, X. Yang, and W. Yu, “Ml-gat: A multilevel graph attention model for stock prediction,” *IEEE Access*, vol. 10, pp. 86 408–86 422, 2022.

[41] N. Wu, B. Green, X. Ben, and S. O’Banion, “Deep transformer models for time series forecasting: The influenza prevalence case,” 2020.

[42] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, pp. 11 106–11 115, 2021.

[43] A. Zeng, M. Chen, L. Zhang, and Q. Xu, “Are transformers effective for time series forecasting?” *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, pp. 11 121–11 128, 2023.

[44] I. Mastromatteo, E. Zarinelli, and M. Marsili, “Reconstruction of financial networks for robust estimation of systemic risk,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, p. P03011, 2012.

[45] I. Zelinka and R. Senkerik, “Chaotic attractors of discrete dynamical systems used in the core of evolutionary algorithms: state of art and perspectives,” *Journal of Difference Equations and Applications*, pp. 1–26, 2023.

Efficient IoT Security: Weighted Voting for BASHLITE and Mirai Attack Detection

Marwan Abu-Zanona
Department of Management Information Systems College of Business Administration
King Faisal University, Al-Ahsa 31982, Saudi Arabia

Abstract—**The increasing number of devices in the Internet of Things (IoT) has exposed various vulnerabilities, such as BASHLITE and Mirai attacks, making it easier for cyber threats to emerge. Due to these vulnerabilities, developing innovative detection and mitigation strategies is essential. Our proposed solution is an ensemble-based weighted voting model that combines different classifiers, including Random Forest, eXtreme Gradient Boosting (XGBoost), Gradient Boosting, K-nearest neighbor (KNN), Multilayer Perceptron (MLP), and Adaptive Boosting (AdaBoost), using artificial intelligence and machine learning. We evaluated our model on the N-BaIoT dataset, a benchmark in this domain. Our results show that the weighted voting approach has exceptional accuracy, precision, recall, and F1-Score. This highlights the effectiveness of our model in classifying various attack instances within the IoT security context. Our approach performs better than other state-of-the-art methods, achieving a remarkable accuracy of 99.9955% in detecting and preventing BASHLITE and Mirai cyber-attacks on IoT devices.**

Keywords*—*Internet of Things; IoT security; BASHLITE; Mirai attacks; ensemble learning

I. Introduction

IoT devices have transformed our daily lives and work [1]. However, the rapid increase in IoT devices has also exposed serious vulnerabilities in the IoT ecosystem [2]. With a vast attack surface, IoT networks have become prime targets for cybercriminals who seek to compromise devices and launch large-scale attacks [3]. Addressing these security concerns and understanding the potential effects of IoT attacks on vital infrastructure, data privacy, and overall societal well-being is essential [4]. IoT devices are vulnerable to numerous security threats due to their constrained resources, diverse communication protocols, and heterogeneous architectures [5]. These vulnerabilities include weak or default credentials, inadequate mechanisms, and the absence of timely software updates [6]. The complexity of security management is further compounded by the heterogeneous nature of IoT networks that comprise devices from various manufacturers [7]. IoT systems face many threats, while the BASHLITE and Mirai attacks are among the most widespread. These attacks have been responsible for numerous large-scale Distributed Denial of Service (DDoS) incidents [8]. The BASHLITE and Mirai attacks have significantly impacted the IoT security landscape, posing a severe threat to connected devices and networks [9]. These cause financial losses by creating powerful botnets. These attacks have become more sophisticated, with malware and botnet operators evolving [10]. Organizations have implemented attack triage systems to combat these threats that help security operators identify and analyze new attack patterns [11]. Analysis of the behavior of these botnets reveals that they rely on infrastructure providers and target specific victims [12]. By understanding the characteristics of these attacks, we can gain valuable insights into the challenges that require innovative detection and mitigation approaches. To effectively secure IoT systems, it is necessary to quickly detect and prevent threats such as BASHLITE and Mirai attacks. Traditional security methods often fail in IoT environments because of the specific features of these devices, such as limited computational resources and constrained communication capabilities [13]. Therefore, efficient detection methods are essential for IoT security. Artificial intelligence and machine learning ensembles can create a more resilient defense against IoT-related threats [14]. Innovative approaches, such as ensemble-based weighted voting, should be adopted to improve detection accuracy and efficiency [15]. Ensemble-based voting optimizes predictive outcomes by fusing diverse algorithms to mitigate inherent uncertainties and strengthen model robustness [16].

Based on the details presented, we will apply the three research questions:

1) How can we effectively detect and prevent BASHLITE and Mirai attacks to enhance the security of IoT systems?
2) What strategies keep up with the changing landscape of BASHLITE and Mirai attacks on IoT networks?
3) How can we improve the accuracy and efficiency of IoT security systems in the face of unique device constraints by applying ensemble-based weighted voting and other artificial intelligence techniques?

The objective of this work is to detect and prevent BASHLITE and Mirai attacks. To achieve this, the researchers developed an ensemble-based weighted voting model that combines machine learning classifiers such as Random Forest, eXtreme Gradient Boosting (XGBoost), Gradient Boosting, K-nearest neighbor (KNN), Multilayer Perceptron (MLP), and Adaptive Boosting (AdaBoost). The model is evaluated through extensive testing on the N-BaIoT dataset and shows different accuracy measurements in identifying attacks in IoT environments. Then compares contemporary techniques and establishes the proposed approach as a state-of-the-art solution. The research focuses on IoT security, explicitly targeting the detection and classification of BASHLITE and Mirai attacks. The key contributions are:

- Ensemble-based weighted voting approach, leveraging machine learning classifiers (Random Forest, XGBoost, KNN, MLP, and AdaBoost).

- Validation on the N-BaIoT Dataset showcased the model's heightened performance in accurately identifying attacks, demonstrating improved accuracy, precision, recall, and F1-Score.
- Comparative analysis of the proposed approach with contemporary techniques, establishing its prowess as a state-of-the-art solution for IoT security.
- The findings provide valuable insights into strengthening IoT networks, showing practical implications for enhancing security measures against growing cyber threats.

The rest of this work is organized as follows: Section II delves into existing related works in the field. Section III details the methodology, covering the N-BaIoT dataset, defining the single classifiers, and defining performance metrics. Subsequently, Section IV details the proposed weighted voting approach, the training process, and the specifics of the ensemble weighted voting technique. The evaluation and results derived from the experimental setup are expounded in Section V. Section VI concludes findings and outlines future research efforts.

II. Related Work

With the increasing threats to IoT networks, there has been a noteworthy rise in research attempts towards IoT security. This section comprehensively reviews the relevant literature on IoT security and attack detection. We analyze the previous studies and approaches that have addressed the challenges IoT devices pose and their vulnerabilities. By synthesizing and analyzing the existing body of work, we aim to gain a deeper understanding of the current state of IoT security research. Ensemble learning techniques have been proposed to identify and classify attacks on IoT networks [17], [18]. These techniques implement machine learning to enhance the detection of security breaches and recommend appropriate mitigation strategies [19]. Ensemble models trained on realistic data have shown promising results in accuracy [20]. Additionally, efficient and lightweight machine learning-based detection systems have been developed to counter attacks on IoT devices [21]. Ensemble learning approaches offer potential solutions for improving attack detection and securing IoT networks.

Alothman et al. [22] proposed an approach using machine learning to detect IoT botnet attacks. This approach was proposed to differentiate malicious traffic from normal traffic and identify botnet types. They utilized the Bot-IoT dataset containing various attack categories and applied preprocessing techniques like Synthetic Minority Over-sampling Technique (SMOTE). Using the Bot-IoT dataset, they tested multiple classifiers and reported the results of the best three classifiers, J48, Random Forest, and MLP. The results showed that the RF and J48 classifiers had superior accuracies of 0.960 and 0.963, respectively.

Alkahtani et al. [23] proposed a hybrid deep learning algorithm to detect botnet attacks on IoT networks. The experimental results showed that the proposed model using the N-BaIoT dataset, achieved near 90% accuracy in detecting attacks from doorbells. For thermostat devices, the proposed system achieved an accuracy of 88.53% in identifying botnet attacks. The proposed system also exhibited high accuracies in detecting botnet attacks from security cameras, achieving from 87.19% to 89.64%.

Karanja et al. [24] designed a method for identity malware in the IoT using Haralick image texture features and three machine learning classifiers. They converted the data to gray scale images and computed the gray-level co-occurrence matrix (GLCM) for each image. Then, they calculated five Haralick features that were used to classify malware using classifiers. The experimental results showed that their approach achieved three results 80%, 89%, and 95% accuracy based on their findings.

Alsamiri et al. [25] used different machine learning algorithms to quickly and effectively detect IoT network attacks. They utilized the Bot-IoT dataset for the evaluation process. During the implementation phase, several algorithms were applied, and most achieved high performance. Additionally, new features were extracted from the Bot-IoT dataset which resulted in better results. Based on the experimental results, the KNN algorithm was found to be the most effective with 99% accuracy.

A research study [26] produced a MedBIoT dataset containing both typical and botnet traffic in the IoT network. The dataset includes data from the primal phase of botnet preparation and features real botnet malware such as Mirai, BASHLITE, and Torii. Machine learning models, both supervised and unsupervised, were built using the data to demonstrate the effectiveness of machine learning-based botnet classification and intrusion detection systems. The experimental results showed that the RF algorithm was the most effective, with an accuracy of 96.17%. The collected dataset has proven suitable and reliable for botnet detection using machine learning techniques.

In [27], the authors proposed new algorithms designed to encrypt data streams in an efficient IoT environment that meets the security requirements of 5G networks. They demonstrated that their algorithms resist various types of attacks, including quantum attacks, eavesdropping, plaintext attacks, chosen ciphertext attacks, and public critical attacks. The authors compared their proposed algorithm with leading post-quantum (PQ) cryptography algorithms such as LWE, LIZARD, and NTRU. According to their findings, the symmetric algorithm they proposed is 70 times faster than the aforementioned symmetric algorithms, and their asymmetric algorithm is ten times faster than the above-stated asymmetric algorithms. Additionally, both the proposed algorithms require 6000 times less memory.

When it comes to IoT devices, we face a challenge: they tend to need more energy and power. Post-quantum cryptography is usually more computationally intensive than the current cryptographic standards. In study [28], authors used the post-quantum digital signature scheme CRYSTALS-Dilithium to authenticate Message Queue Telemetry Transport (MQTT) and measure CPU, memory, and disk usage. They also explored using a key encapsulation mechanism (KEM) trick suggested in 2020 for transport-level security (TLS), which can save up to 90% of CPU cycles. They utilized the post-quantum KEM scheme CRYSTALS-KYBER to compare the resulting CPU, memory, and disk usage with traditional authentication. The study found that using KEM for authentication resulted in a 25 ms speed increase and a 71% savings. Although there were some additional costs for memory, they were minimal enough to be acceptable for most IoT devices.

Some maintain the trade-offs among cost, performance, and security, especially when considering resource-constrained IoT devices. The authors in [29] discussed various S-boxes used in the popular LWC algorithms by their input–output bit-sizes and highlighted their strengths and limitations. Then, it focuses on the proposed 5-bit S-box design. The novel design uses a chaotic mapping theory to offer a random behavior of the element in the proposed S-box. The experimental results from ASIC implementation reveal two essential characteristics of the proposed S-box, cost and performance, and further compare it with 4/5-bit competitors. It demonstrates the security strength of the proposed 5-bit S-box through cryptanalyses such as bijective, nonlinearity, linearity, differential cryptanalysis, differential style boomerang attack, avalanche effect, bit and independence criterion. Also, a comparison is carried out to exhibit the superiority of the proposed 5-bit S-box over its 5-bit competitors.

The article in [1] presented an efficient design method for PSCA-resistant ciphers implemented in hardware using high-level synthesis. The focus is on lightweight block ciphers that use addition, rotation, and XOR-based permutations. They also studied the effects of threshold implementation, which is a secure countermeasure against power side-channel attacks, on the behavioral descriptions of ciphers, along with changes in high-level synthesis scheduling. The proposed method successfully improves the resistance against power side-channel attacks for all addition/rotation/XOR-based ciphers used as benchmarks, as demonstrated by the results obtained using Welch's t-test.

III. Methodology

A. The N-BaIoT Dataset

The N-BaIoT dataset, introduced by Meidan et al. [30], comprises 115 attributes obtained through port mirroring of IoT devices. It includes benign data captured immediately after network setup, featuring two types of packet sizes, packet counts, and jitters. These data samples are categorized based on source IP, source MAC-IP, channel, socket, and total, with attributes such as packet, packet count, and time between packet arrivals. Statistical measures like mean, variance, integer values, magnitude, radius, covariance, and correlation coefficient are covered across 23 features for each of the five-time windows. The dataset incorporates injected BASHLITE and Mirai attacks into various IoT devices, each associated with specific device types and model names. BASHLITE, also known as gafgyt, is a botnet for DDoS attacks on Linux-based IoT devices, employing flooding attacks like UDP and TCP attacks. In contrast, Mirai, developed by Paras, executes large-scale attacks on IoT devices by scanning for vulnerabilities and launching flooding attacks. This dataset offers a comprehensive understanding of IoT device behavior under benign and malicious conditions, providing insights into the impact of different attacks on various device types.

B. Single Classifiers

Our ensemble model leverages the unique features of each classifier. Each classifier has a distinct architecture, hyperparameters, and beneficial features for IoT security. Our approach employs various machine learning models, including the Random Forest Classifier, XGBoost, Gradient Boosting Classifier, KNN Classifier, MLP Classifier, and AdaBoost Classifier, to classify attacks and anomalies in the proposed data. Every model plays an essential role in our approach to enhancing attack detection in IoT ecosystems.

1) Random Forest Classifier: The Random Forest Classifier is a machine learning algorithm for classification and regression tasks [31]. It is a reliable and robust technique that enhances predictive accuracy by utilizing multiple decision trees while minimizing overfitting. The algorithm creates a collection of decision trees using bootstrapping to build each tree from a subset of the dataset. During the creation of each tree, a random subset of features is considered for splitting at each node. This diversity is necessary to ensure that the forest is not overly dependent on a particular set of features, promoting robustness and flexibility in the predictive system [32]. The prediction process involves aggregating the outputs of individual trees, where each tree casts a vote for the class label. Overall, the Random Forest Classifier is a powerful tool for machine learning tasks that require high accuracy and flexibility. Let T denote the set of decision trees in the forest, and $h_i(\mathbf{x})$ represent the prediction of the i-th tree for input vector $\mathbf{x}$. The final prediction $\hat{y}$ is determined by a majority vote:

$$\hat{y} = \underset{y}{\operatorname{argmax}} \left(\sum_{i=1}^{|T|} \mathbb{I}\{h_i(\mathbf{x}) = y\} \right) \tag{1}$$

where, $\mathbb{I}\{\cdot\}$ is the indicator function, and $|T|$ signifies the total number of trees in the Random Forest. This approach allows Random Forest to make accurate predictions by combining decisions from multiple trees, improving adaptability, and reducing overfitting.

2) Gradient Boosting Classifier: The Gradient Boosting Classifier is an advanced machine learning algorithm that constructs predictive models sequentially [33]. It achieves this by addressing the errors of its predecessors, which refine the model's accuracy and make it particularly effective for classification tasks. The core idea behind this algorithm is to combine the outputs of multiple weak learners, usually decision trees, in a weighted manner [34]. Each tree is trained to rectify the mistakes of the preceding ones, optimizing the overall predictive performance. The algorithm employs a scheme that minimizes a composite objective function. It contains a loss term that quantifies prediction errors and regularization terms for controlling model complexity. Through gradient descent, the classifier adjusts the parameters of each weak learner to improve its predictive capabilities. This approach enables Gradient Boosting to excel at capturing intricate patterns within data, making it a valuable tool for various real-world applications. In the Gradient Boosting Classifier, the prediction is formulated as an additive ensemble of weak learners, typically decision trees. The overall prediction $F(x)$ is a weighted sum of these weak learners, given by the equation:

$$F(x) = \sum_{t=1}^{T} \alpha_t f_t(x) \tag{2}$$

where, T represents the total number of weak learners in the ensemble, α_t denotes the weight assigned to the t-th weak learner, and $f_t(x)$ is the prediction made by the t-th weak learner. Gradient boosting works by adding weak learners sequentially. Each new learner addresses the residuals (errors) of the combined model from the previous iterations. This approach helps in improving the overall model accuracy gradually by focusing on the previously misclassified instances. During the training process, the weights (α_t) are determined, which depend on the contribution of each weak learner to minimizing the overall loss function.

3) XGBoost Classifier: XGBoost is a highly efficient and versatile ensemble learning algorithm that is widely used for predictive modeling [35]. It is a gradient boosting method that builds a series of weak learners, such as decision trees, and adaptively improves their predictive performance. XGBoost includes regularization terms in its objective function, which promotes model simplicity and reduces overfitting [36]. XGBoost optimizes predictive accuracy with its innovative approach to tree construction and parallel processing. The XGBoost classifier algorithm aims to improve the predictive capabilities of an ensemble of weak learners, typically decision trees. It does this by minimizing the objective function through an iterative boosting process that adjusts the parameters of each weak learner. The final prediction is determined by aggregating the weighted contributions of individual learners, where weights reflect the influence of each learner on the overall model. The success of XGBoost in achieving superior performance across diverse classification scenarios is due to its gradient descent optimization and regularization. The XGBoost classifier is defined as:

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i) \tag{3}$$

While, $\hat{y}_i$ represents the predicted output for the i-th instance, $\mathbf{x}_i$ denotes the feature vector, K is the total number of weak learners, and $f_k(\mathbf{x}_i)$ represents the output of the k-th weak learner. The ensemble prediction is obtained by summing the outputs of all weak learners, and $\phi(\mathbf{x}_i)$ produces the final predicted value. The parameters of the weak learners are adaptively updated during the iterative training process, allowing the model to learn complex relationships in the data.

4) K-Nearest Neighbors (KNN): The KNN classifier is a type of machine learning algorithm used for classification and regression tasks that does not require any pre-defined parameters [37]. It works by determining the classification of a data point based on the majority class of its KNNs in the feature space. The algorithm often uses Euclidean distance as a metric to measure similarity between data points. The consensus of the classes within the neighborhood of a point forms the decision boundary. The KNN algorithm is versatile and can adapt to different data and decision boundaries. However, choosing the appropriate value for k is essential since a small value of k can increase sensitivity to noise, while a high value of k may smooth out local patterns. Despite its simplicity, KNN has proven effective in various applications, especially when decision boundaries are intricate and non-linear [38]. The KNN classifier predicts the class label of a data point by considering the majority class among its k-nearest neighbors in the feature space [39]. Let x be the data point for which we want to make a prediction, and let $N(x)$ denote the set of its k-nearest neighbors based on a specified distance metric. The predicted class label, $\hat{y}$, is determined by the majority class among these neighbors. It is presented in the equation:

$$\hat{y} = \arg\max_{y} \left(\sum_{i=1}^{k} \mathbb{I}\{y_i = y\} \right) \tag{4}$$

where, y_i represents the class labels of the k-nearest neighbors. The decision is made by selecting the class y that maximizes the count of its occurrences among the neighbors. This formulation captures the essence of the KNN algorithm, emphasizing the reliance on local neighborhood information for classification decisions. The choice of the distance metric and the parameter k significantly influences the algorithm's performance and its adaptability to different datasets.

5) Multi-Layer Perceptron (MLP): The MLP classifier is an artificial neural network comprising several layers of interconnected nodes, each serving as a processing unit [40]. It is a versatile and powerful model that can be used for both classification and regression tasks. In classification, the MLP classifier uses a feed-forward architecture, where information moves from the input layer through hidden layers to the output layer. The hidden layers have nodes that use non-linear activation functions, enabling the model to capture complex relationships in the data. Overall, the MLP classifier is a reliable tool for handling complex datasets [41]. Let $\mathbf{x}$ represent the input vector, and W and b denote the weight matrix and bias vector, respectively, for each layer. The output of the MLP is obtained through a series of transformations and activation functions. The prediction $\hat{y}$ is calculated as:

$$\hat{y} = \sigma(W_{\text{out}} \cdot \sigma(W_{\text{hidden}} \cdot \sigma(W_{\text{input}} \cdot \mathbf{x} + b_{\text{input}}) + b_{\text{hidden}}) + b_{\text{out}} \tag{5}$$

σ denotes the activation function, such as the Rectified Linear Unit (ReLU) or hyperbolic tangent (tanh). During training, the weights and biases of MLPs are optimized using back propagation and gradient descent algorithms. The flexibility of these models enables them to learn complex patterns and relationships present in the data, which makes them an excellent fit for a wide range of machine learning applications.

C. Performance Metrics

This section establishes a set of evaluation metrics to assess the performance of our ensemble model and its components quantitatively. We have chosen metrics that help evaluate accuracy, precision, recall, and F1-score. These metrics provide objective benchmarks for comparing our approach with existing detection methods. Additionally, we discuss our methodology for conducting experiments and collecting results. Selecting and interpreting evaluation metrics is essential to ensure a comprehensive and informative assessment of our research outcomes. We use several standard evaluation metrics for accuracy, precision, recall, and F1-score.

- **Accuracy (ACC):** Accuracy is a metric that measures the ratio of correctly predicted instances to the total instances in the dataset. ACC is calculated as,

$$\text{ACC} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}} \tag{6}$$

- **Precision (P):** Precision measures the model's ability to identify true positives accurately. P is calculated as,
$$\text{P} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{7}$$
- **Recall (R):** Recall is a metric that indicates the percentage of actual positive cases that the model correctly predicted. R is calculated as,
$$\text{R} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{8}$$
- **F1-Score (F1):** The F1-score provides a balance between precision and recall. F1 is calculated as,
$$\text{F1} = \frac{2 \cdot \text{P} \cdot \text{R}}{\text{P} + \text{R}} \tag{9}$$

These metrics are essential to objectively evaluate the performance of our ensemble model in detecting BASHLITE and Mirai attacks. They enable comparisons with other detection methods and offer a comprehensive assessment of our research outcomes.

IV. Proposed Weighted Voting Approach

A. Proposed Approach Overview

The proposed approach involves a weighted voting strategy using an ensemble of five single classifiers (as described in Section III-B). These classifiers are trained with a prepared dataset and their predictions are assigned weights based on their performance and confidence. Fig. 1 provides a clear overview of our proposed approach.

B. Preprocessing

In this step, we will discuss data collection, and preprocessing procedures, namely feature encoding and scaling. Feature encoding involves transforming categorical variables into a numerical format, which helps machine learning algorithms to understand and represent categorical data accurately. Feature scaling is an essential technique that plays a pivotal role in normalizing the range of numerical features. The process helps to prevent any feature from disproportionately influencing the learning process. Feature scaling is an essential technique in machine learning, and the Standard Scaler is one of the most commonly used methods [42]. Its primary purpose is to normalize the range of numerical features within a dataset. This helps to prevent any feature from having an outsized impact on the learning process. By transforming the data distribution to have a mean of 0 and a standard deviation of 1, the Standard Scaler ensures all features are brought to a common scale. For each feature x_i in the dataset, the Standard Scaler computes the z-score by subtracting the mean (μ) of the feature and dividing it by its standard deviation (σ), as expressed in the equation:
$$z = \frac{(x_i - \mu)}{\sigma} \tag{10}$$
Here, z means the standardised set of the characteristic x_i. This process ensures the stability and effectiveness of the training process in interpreting and utilizing numerical features. By meticulously scaling our features, we can improve the performance and reliability of our model in the later stages of our research.

C. Training Machine Learning Classifiers

The training data included labels that explained whether a particular output had an anticipated associated class. The main goal is to train the learning model to perceive the correct position of the unseen data by matching it to the sample data. However, in many cases, we found that a single learning model could have produced the best results or the minimum errors. Therefore, we adopted an ensemble learning technique that involved constructing multiple assumptions on the training data and incorporating them to recognize the correct position of the sample. This method combined the decisions from several models and enhanced the overall efficiency of the model, resulting in more accurate results. Moreover, this approach led to a stable and more robust model than individual models.

To prepare our ensemble model, we meticulously execute the training process for each machine learning classifier making up our ensemble. The classifiers as described in Section III-B, include the Random Forest, XGBoost, Gradient Boosting, KNN, MLP, and AdaBoost Classifiers. Each classifier's diverse architectures, hyperparameters, and unique capabilities contribute to the comprehensive learning process. Subsequently, these trained classifiers collectively form the basis learned for the ensemble approach, paving the way for the ensemble's ultimate strength through Weighted Voting in detecting malicious activities within IoT environments.

D. Ensemble Weighted Voting

The proposed approach relies on the Weighted Voting strategy, which allows us to use the unique strengths of individual models within the ensemble [43]. It delves into the intricacies of Weighted Voting, explaining how it assigns weights to predictions from base trained models based on their performance and confidence [44]. This approach aims to optimize the accuracy of our ensemble system, enabling it to adapt to varying degrees of model reliability. Leveraging these weighted predictions is essential in improving the detection of BASHLITE and Mirai attacks in IoT networks. Let C_i represent the i-th base classifier, where i ranges from 1 to n (n = 5). Each base classifier provides a prediction denoted as P_i, and these predictions collectively form the set $\{P_1, P_2, ..., P_n\}$. Next, individual weights are assigned to these predictions based on the performance or reliability of each base classifier. Let W_i denote the weight assigned to the prediction of classifier C_i. The assignment of weights can be influenced by various factors, such as the accuracy or precision of each base classifier on a validation set. The weighted sum, denoted as S, is computed by summing the product of each prediction and its corresponding weight:
$$S = \sum_{i=1}^{n} W_i \cdot P_i \tag{11}$$
In ensemble learning, the final prediction output is the sum of the five predictions made by all the base classifiers. Each base classifier's prediction has a weight assigned to it, and the final output is a combination of all these weighted predictions. To ensure that the weights form a proper distribution, they are often normalized. This means that each weight is divided by

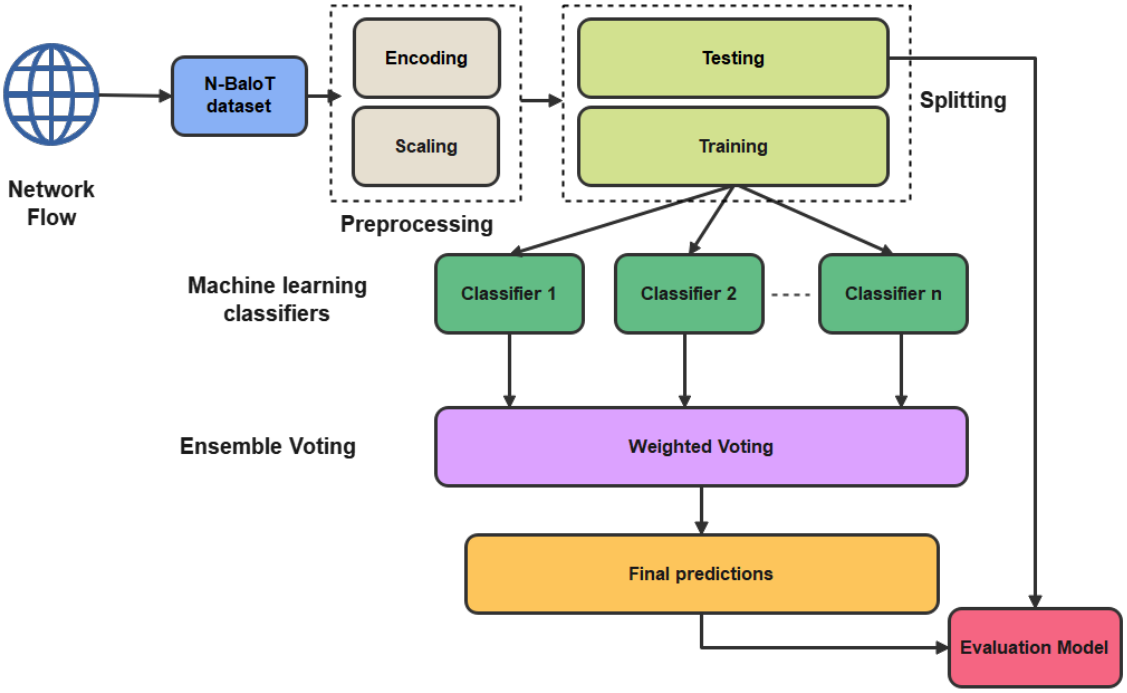


Fig. 1. Flow diagram of the proposed weighted voting attack detection approach.

the sum of all the weights:

$$\text{Normalized Weight, } \bar{W}_i = \frac{W_i}{\sum_{j=1}^{n} W_j} \tag{12}$$

This normalization guarantees that the weights collectively sum to 1, maintaining the integrity of the weighted voting process. The Weighted Voting mechanism ensures that the final prediction is a well-informed combination of individual base classifier predictions. Each contribution is appropriately weighted to enhance the overall accuracy of the ensemble. The mechanism's effectiveness lies in its ability to assign weights adaptively based on the reliability of each classifier. This helps to improve the predictive accuracy of the ensemble.

V. Evaluation and Results

In this section, we thoroughly evaluate and present the results of our proposed ensemble approach for detecting BASHLITE and Mirai attacks from the N-BaIoT dataset in IoT networks. Our experimental evaluation meticulously analyzes the model's performance using key metrics such as accuracy, precision, recall, and F1-score. We provide insightful analyses of the experimental setup and the obtained results and then conduct a comparative assessment with other existing detection methods. This comprehensive evaluation presents a detailed perspective on the strengths of our ensemble model and potential enhancements in IoT security.

A. Experimental Setup

TABLE I. Parameters and Methods for the Machine Learning Models

Model	**The parameters used for the model experiments.**
Random Forest	- **Number of Trees :** 100. - **Criterion for Splitting (criterion):** 'gini'. Methods:'entropy'.
Gradient Boosting	- **Number of Boosting Stages (n_estimators):** 100. Methods: Tune based on the trade-off between performance and computational cost. - **Loss Function (loss):** 'deviance'. Methods:'exponential'. - **Learning Rate:** 0.1.
K Nearst Neighbors	- **Number of Neighbors (n_neighbors):** 5. Methods: Square root of the number of samples. - **Weight Function (weights):** 'uniform'.
MLP	- **Number of Neurons in Hidden Layers (hidden_layer_sizes):** 100.- **Activation Function (activation):** 'relu'. Methods: 'tanh'.
AdaBoost	- **Number of Weak Learners (n_estimators):** 50. - **Learning Rate (learning_rate):** 1.0.
XGBoost	- **Number of Trees (n_estimators):** 100.

In this section, we explain the experimental settings of our contribution designed for the detection of BASHLITE and Mirai attacks in IoT networks. We provide a detailed account of the implementation steps and propose solutions aimed at validating the effectiveness of our approach in enhancing the security of IoT environments. A critical aspect of the validation process involves the assessment of the N-BaIoT dataset, underscoring its significance in evaluating any proposed solution for IoT security improvement. Table I summarizes parameters and methods employed for various machine learning models, explaining the comprehensive experimental setup to strengthen IoT security. We split the N-BaIoT dataset into two sets-training and testing, with 80% for training and 20% for model evaluation. This helped balance model learning and validation, measuring our ensemble model's real-world performance while minimizing overfitting.

B. Experimental Results

The study employed Python on Google Colab GPU for multiclass classification [45]. To detect BASHLITE and Mirai attacks, we employed five machine learning techniques (Random Forest, XGBoost, Gradient Boosting, KNN, MLP, and AdaBoost). The results were combined using a weighted voting technique. The ensemble approach and the five classifiers were evaluated using performance evaluation measures such as accuracy, precision, recall, and F1-score. The performance of different machine learning classifiers was compared, including the Ensemble Weighted Voting approach and individual classifiers. The results presented in Tables II, and III, show the weighted and macro average performance metrics.

TABLE II. WEIGHTED AVERAGED FOR DIFFERENT METRICS ACROSS MODELS (IN %)

Model	Accuracy	Precision	Recall	F1-Score
Ensemble Weighted Voting	99.9955	99.9955	99.9913	99.9955
XGBoosting	99.9796	99.9796	99.9624	99.9796
KNN	99.9087	99.9087	99.8669	99.9087
Random Forest	99.9909	99.9909	99.9847	99.9909
MLP	99.9781	99.9781	99.9663	99.9781
AdaBoost	69.13	65.14	59.33	65.14

TABLE III. MACRO AVERAGED METRICS FOR DIFFERENT MODELS (IN %)

Model	Precision	Recall	F1-Score
Ensemble Weighted Voting	99.99	99.99	99.99
XGBoosting	99.97	99.96	99.96
KNN	99.87	99.86	99.87
Random Forest	99.98	99.98	99.98
MLP	99.97	99.96	99.97
AdaBoost	66.20	66.20	66.20

The Ensemble Weighted Voting model outperformed the individual classifiers across various metrics such as accuracy, precision, recall, and F1-score. The Ensemble Weighted Voting model has shown exceptional accuracy in the Weighted Averaged metrics, achieving a score of 99.9955%. This is higher than all other classifiers, including XGBoosting, KNN, Random Forest, MLP, and AdaBoost. Additionally, the model's precision, recall, and F1-Score are consistently superior, reaching 99.9955%, 99.9913%, and 99.9955%, respectively. Its higher precision and recall values indicate a better ability to correctly identify and classify instances of attacks, leading to a high F1-score. In the macro-averaged metrics, which measure the average performance across different classes, the Ensemble Weighted Voting model outperforms its competitors with precision, recall, and F1-Score values of 99.99%, 99.99%, and 99.99%, respectively. On the other hand, the individual classifiers, including AdaBoost, exhibited lower precision, recall, and F1-score performance. The ensemble's weighted average results highlight its effectiveness in combining diverse models, emphasizing the importance of ensemble learning in achieving improved detection capabilities in IoT security applications. This indicates that the model is robust and reliable, with a superior ability to detect various classes within the IoT security context.

C. Discussion

This section discusses the implications of the study's findings and the recent techniques used to detect and prevent BASHLITE and Mirai cyber-attacks on IoT devices based on evaluating the N-BaIoT dataset. Our proposed solution focused on an ensemble-based weighted voting model, representing a comprehensive IoT security approach. This model combines various classifiers, including Random Forest, XGBoost, Gradient Boosting, KNN, MLP, and AdaBoost, utilizing artificial intelligence and machine learning techniques. The evaluation of the N-BaIoT dataset, a recognized benchmark in the domain, demonstrates the exceptional performance of the weighted voting approach. The model achieves outstanding accuracy, precision, recall, and F1-Score, surpassing state-of-the-art methods. Notably, the accuracy of 99.9955% in detecting and preventing BASHLITE and Mirai cyber-attacks on IoT devices is a remarkable highlight. Comparing our approach with recent works, as summarized in Table IV, affirms the superiority of the proposed model. Abu Al-Haija and Al-Dala'ien [46] used different machine learning techniques AdaBoosted, RUSBoosted, and bagged, achieving a detection accuracy of 99.6% for botnet attacks.

TABLE IV. COMPARISON BETWEEN MOST RECENT WORKS AND THE PROPOSED APPROACH WITH RESULTS FINDINGS

Authors	Dataset	Results in %
Ours	N-BaIoT	99.9955
Abu Al-Haija and Al-Dala'ien [46]	N-BaIoT	99.6
Okur et al. [47]	N-BaIoT	99.92
Sakthipriya et al. [48]	N-BaIoT	95.02
Abbasi et al. [49]	N-BaIoT	above 90
Hezam et al. [50]	N-BaIoT	89.75

Okur et al. [47] reported a 99.92% accuracy detection rate using Random Forest in the N-BaIoT dataset. These results provide a context for understanding the competitive edge of our ensemble-based approach. Furthermore, examining alternative methods, Sakthipriya et al. [48] focused on dimensionality reduction, with auto-encoder outperforming PCA with an accuracy of 95.02%. Abbasi et al. [49] proposed logistic regression for intrusion detection, achieving above 90% classification accuracy, while Hezam et al. [50] explored deep learning algorithms, with RNN achieving the highest accuracy of 89.75%. The comparison presented in Table IV highlights the superior accuracy of our proposed model compared to recent uses the N-BaIoT dataset. Our model's performance was a comprehensive ensemble of classifiers used to prevent IoT security issues with robustness and effectiveness. The evaluation process was meticulous, ensuring the accuracy of the results.

VI. CONCLUSION AND FUTURE SCOPE

The increasing number of devices in the IoT has also led to an increase in security risks, such as BASHLITE and Mirai attacks. To address these vulnerabilities, we need innovative detection and mitigation strategies. In this study, we present a new solution based on an ensemble-based weighted voting model that uses a variety of classifiers, including Random Forest, XGBoost, Gradient Boosting, KNN, MLP, and AdaBoosting, powered by artificial intelligence and machine learning. We rigorously evaluated the model's effectiveness on the N-BaIoT dataset, a benchmark in the IoT security domain. Our results indicate that the proposed weighted voting approach achieves exceptional accuracy, precision, recall, and F1-Score in accurately classifying various attack instances in the IoT security context, outperforming other state-of-the-art methods. Notably, the model demonstrates an outstanding accuracy rate of 99.9955% in detecting and preventing BASHLITE and Mirai cyber-attacks on IoT devices. The proposed ensemble-based weighted voting model is designed to overcome the challenges posed by BASHLITE and Mirai attacks and provides valuable insights into IoT networks. Combining different machine learning classifiers, the model shows superior performance metrics to individual classifiers, making it adaptable to changing attack patterns. This study aims to protect against current IoT security threats, providing a robust defense model.

In the future, we can explore how attack methods evolve and test more machine learning classifiers to see how well the proposed ensemble-based approach adapts. We can also study how well the model performs in real-world IoT environments, and we can improve its practical usefulness by testing how it handles larger datasets. Continuously improving and expanding the model using different artificial intelligence techniques and cybersecurity advancements ensure its effectiveness in the ever-changing landscape of IoT security.

Acknowledgment

The researcher extends his sincere thanks and appreciation to the administration of King Faisal University in the Kingdom of Saudi Arabia for providing all forms of support to the university's faculty members, especially in the field of scientific research. We extend our sincere thanks to the Deanship of Scientific Research at the university for agreeing to support this scientific research with GRANT number 5,290.

References

[1] D. Singh, "Internet of things," *Factories of the Future: Technological Advancements in the Manufacturing Industry*, pp. 195–227, 2023.

[2] S. Ahmed and M. Khan, "Securing the internet of things (iot): A comprehensive study on the intersection of cybersecurity, privacy, and connectivity in the iot ecosystem," *AI, IoT and the Fourth Industrial Revolution Review*, vol. 13, no. 9, pp. 1–17, 2023.

[3] P. K. Sadhu, V. P. Yanambaka, and A. Abdelgawad, "Internet of things: Security and solutions survey," *Sensors*, vol. 22, no. 19, p. 7433, 2022.

[4] M. H. P. Rizi and S. A. H. Seno, "A systematic review of technologies and solutions to improve security and privacy protection of citizens in the smart city," *Internet of Things*, vol. 20, p. 100584, 2022.

[5] B. B. Gupta and M. Quamara, "An overview of internet of things (iot): Architectural aspects, challenges, and protocols," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 21, p. e4946, 2020.

[6] M. D. Iannacone and R. A. Bridges, "Quantifiable & comparable evaluations of cyber defensive capabilities: A survey & novel, unified approach," *Computers & Security*, vol. 96, p. 101907, 2020.

[7] A. E. Omolara, A. Alabdulatif, O. I. Abiodun, M. Alawida, A. Alabdulatif, H. Arshad *et al.*, "The internet of things security: A survey encompassing unexplored areas and new insights," *Computers & Security*, vol. 112, p. 102494, 2022.

[8] P. Kumari and A. K. Jain, "A comprehensive study of ddos attacks over iot network and their countermeasures," *Computers & Security*, p. 103096, 2023.

[9] M. H. Aysa, A. A. Ibrahim, and A. H. Mohammed, "Iot ddos attack detection using machine learning," pp. 1–7, 2020.

[10] A. Marzano, D. Alexander, O. Fonseca, E. Fazzion, C. Hoepers, K. Steding-Jessen, M. H. Chaves, Í. Cunha, D. Guedes, and W. Meira, "The evolution of bashlite and mirai iot botnets," pp. 00 813–00 818, 2018.

[11] S. Coltellese, F. Maria Maggi, A. Marrella, L. Massarelli, and L. Querzoni, "Triage of iot attacks through process mining," pp. 326–344, 2019.

[12] G. Bastos, A. Marzano, O. Fonseca, E. Fazzion, C. Hoepers, K. Steding-Jessen, Í. Cunha, D. Guedes, and W. Meira, "Identifying and characterizing bashlite and mirai c&c servers," pp. 1–6, 2019.

[13] F. Meneghello, M. Calore, D. Zucchetto, M. Polese, and A. Zanella, "Iot: Internet of threats? a survey of practical security vulnerabilities in real iot devices," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8182–8201, 2019.

[14] I. H. Sarker, A. I. Khan, Y. B. Abushark, and F. Alsolami, "Internet of things (iot) security intelligence: a comprehensive overview, machine learning solutions and research directions," *Mobile Networks and Applications*, vol. 28, no. 1, pp. 296–312, 2023.

[15] A. M. Bamhdi, I. Abrar, and F. Masoodi, "An ensemble based approach for effective intrusion detection using majority voting," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 2, pp. 664–671, 2021.

[16] P. Verma, A. R. K. Kowsik, R. Pateriya, N. Bharot, A. Vidyarthi, and D. Gupta, "A stacked ensemble approach to generalize the classifier prediction for the detection of ddos attack in cloud network," *Mobile Networks and Applications*, pp. 1–15, 2023.

[17] M. Mohy-Eddine, A. Guezzaz, S. Benkirane, M. Azrour, and Y. Farhaoui, "An ensemble learning based intrusion detection model for industrial iot security," *Big Data Mining and Analytics*, vol. 6, no. 3, pp. 273–287, 2023.

[18] M. M. Alani and E. Damiani, "Xrecon: An explainbale iot reconnaissance attack detection system based on ensemble learning," *Sensors*, vol. 23, no. 11, p. 5298, 2023.

[19] K. Keserwani, A. Aggarwal, and A. Chauhan, "Attack detection in industrial iot using novel ensemble techniques," pp. 1–6, 2023.

[20] M. Koppula, L. J. LM *et al.*, "Lnkdsea: Machine learning based iot/iiot attack detection method," pp. 655–662, 2023.

[21] N. Pandey and P. K. Mishra, "Detection of ddos attack in iot traffic using ensemble machine learning techniques," *NHM*, vol. 18, no. 3, pp. 1393–1409, 2023.

[22] Z. Alothman, M. Alkasassbeh, and S. Al-Haj Baddar, "An efficient approach to detect iot botnet attacks using machine learning," *Journal of High Speed Networks*, vol. 26, no. 3, pp. 241–254, 2020.

[23] H. Alkahtani and T. H. Aldhyani, "Botnet attack detection by using cnn-lstm model for internet of things applications," *Security and Communication Networks*, vol. 2021, pp. 1–23, 2021.

[24] E. M. Karanja, S. Masupe, and M. G. Jeffrey, "Analysis of internet of things malware using image texture features and machine learning techniques," *Internet of Things*, vol. 9, p. 100153, 2020.

[25] J. Alsamiri and K. Alsubhi, "Internet of things cyber attacks detection using machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 12, 2019.

[26] A. Guerra-Manzanares, J. Medina-Galindo, H. Bahsi, and S. Nõmm, "Using medbiot dataset to build effective machine learning-based iot botnet detection systems," pp. 222–243, 2020.

[27] A. Kaushik, L. S. S. Vadlamani, M. M. Hussain, M. Sahay, R. Singh, A. K. Singh, S. Indu, P. Goswami, and N. G. V. Kousik, "Post quantum public and private key cryptography optimized for iot security," *Wireless Personal Communications*, vol. 129, no. 2, pp. 893–909, 2023.

[28] J. Samandari and C. Gritti, "Post-quantum authentication in the mqtt protocol," *Journal of Cybersecurity and Privacy*, vol. 3, no. 3, pp. 416–434, 2023.

[29] V. A. Thakor, M. A. Razzaque, A. D. Darji, and A. R. Patel, "A novel 5-bit s-box design for lightweight cryptography algorithms," *Journal of Information Security and Applications*, vol. 73, p. 103444, 2023.

[30] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, and Y. Elovici, "N-baiot—network-based detection of iot botnet attacks using deep autoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 12–22, 2018.

[31] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines," *Ore Geology Reviews*, vol. 71, pp. 804–818, 2015.

[32] J. Hatwell, M. M. Gaber, and R. M. A. Azad, "Chirps: Explaining random forest classification," *Artificial Intelligence Review*, vol. 53, pp. 5747–5788, 2020.

[33] O. Alshboul, A. Shehadeh, G. Almasabha, and A. S. Almuflih, "Extreme gradient boosting-based machine learning approach for green building cost prediction," *Sustainability*, vol. 14, no. 11, p. 6651, 2022.

[34] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Applied Soft Computing*, vol. 74, pp. 634–642, 2019.

[35] S. S. Dhaliwal, A.-A. Nahid, and R. Abbas, "Effective intrusion detection system using xgboost," *Information*, vol. 9, no. 7, p. 149, 2018.

[36] K. Budholiya, S. K. Shrivastava, and V. Sharma, "An optimized xgboost based diagnostic system for effective prediction of heart disease," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 7, pp. 4514–4523, 2022.

[37] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.

[38] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," pp. 1255–1260, 2019.

[39] A. X. Wang, S. S. Chukova, and B. P. Nguyen, "Ensemble k-nearest neighbors based on centroid displacement," *Information Sciences*, vol. 629, pp. 313–323, 2023.

[40] R. Qaddoura, A. M. Al-Zoubi, H. Faris, and I. Almomani, "A multi-layer classification approach for intrusion detection in iot networks based on deep learning," *Sensors*, vol. 21, no. 9, p. 2987, 2021.

[41] J. Naskath, G. Sivakamasundari, and A. A. S. Begum, "A study on different deep learning algorithms used in deep neural nets: Mlp som and dbn," *Wireless Personal Communications*, vol. 128, no. 4, pp. 2913–2936, 2023.

[42] M. M. Ahsan, M. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of data scaling methods on machine learning algorithms and model performance," *Technologies*, vol. 9, no. 3, p. 52, 2021.

[43] A. Dogan and D. Birant, "A weighted majority voting ensemble approach for classification," pp. 1–6, 2019.

[44] A. Nazir, J. He, N. Zhu, A. Wajahat, X. Ma, F. Ullah, S. Qureshi, and M. S. Pathan, "Advancing iot security: A systematic review of machine learning approaches for the detection of iot botnets," *Journal of King Saud University-Computer and Information Sciences*, p. 101820, 2023.

[45] M. Canesche, L. Bragança, O. P. V. Neto, J. A. Nacif, and R. Ferreira, "Google colab cad4u: Hands-on cloud laboratories for digital design," pp. 1–5, 2021.

[46] Q. Abu Al-Haija and M. Al-Dala'ien, "Elba-iot: an ensemble learning model for botnet attack detection in iot networks," *Journal of Sensor and Actuator Networks*, vol. 11, no. 1, p. 18, 2022.

[47] C. Okur, A. Orman, and M. Dener, "Ddos intrusion detection with machine learning models: N-baiot data set," pp. 607–619, 2022.

[48] N. Sakthipriya, V. Govindasamy, and V. Akila, "A comparative analysis of various dimensionality reduction techniques on n-baiot dataset for iot botnet detection," pp. 1–6, 2023.

[49] F. Abbasi, M. Naderan, and S. E. Alavi, "Intrusion detection in iot with logistic regression and artificial neural network: further investigations on n-baiot dataset devices," *Journal of Computing and Security*, vol. 8, no. 2, pp. 27–42, 2021.

[50] A. A. Hezam, S. A. Mostafa, A. A. Ramli, H. Mahdin, and B. A. Khalaf, "Deep learning approach for detecting botnet attacks in iot environment of multiple and heterogeneous sensors," pp. 317–328, 2021.

Sentiment Analysis on Banking Feedback and News Data using Synonyms and Antonyms

Aniruddha Mohanty, Ravindranath C. Cherukuri
Department of Computer Science and Engineering
CHRIST (DEEMED TO BE UNIVERSITY)
Bangalore, India 560074

***Abstract*—Sentiment analysis is crucial for deciphering customers' enthusiasm, frustration, and the market mood within the banking sector. This importance arises from financial data's specialized and sensitive nature, enabling a deeper understanding of customer sentiments. In today's digital and social marketing landscape within the banking and financial sector, sentiment analysis is significant in shaping customer insights, product development, brand reputation management, risk management, customer service improvement, fraud detection, market research, compliance regulations, etc. This paper introduces a novel approach to sentiment analysis in the banking sector, emphasizing integrating diverse text features to enable dynamic analysis. This proposed approach aims to assess the sentiment score of distinct words used within a document and classify them as positive, negative, or neutral. After rephrasing sentences using synonyms and antonyms of unique words, the system calculates sentence similarity using a distance control mechanism. Then, the system updates the dataset with the positive, negative, and neutral labels. Ultimately, the ELECTRA model utilizes the self-trained sentiment-scored data dictionary, and the newly created dataset is processed using the SoftMax activation function in combination with a customized ADAM optimizer. The approach's effectiveness is confirmed through the analysis of post-bank customer feedback and the phrase bank dataset, yielding accuracy scores of 92.15% and 93.47%, respectively. This study stands out due to its unique approach, which centers on evaluating customer satisfaction and market sentiment by utilizing sentiment scores of words and assessing sentence similarities.**

Keywords*—*ELECTRA; Synonyms and antonyms; sentiment analysis; datasets; sentiment score; control distance

I. Introduction

Enhanced comprehension of customer perceptions regarding various banking products and services entails evaluating customers' sentiments, opinions, and attitudes [1]. In the age of digitalization and advanced data analytics, the analysis of social media feedback data has become prevalent. The Banking, Financial Services, and Insurance (BFSI) sector utilizes Customer Relationship Management (CRM) to make informed business decisions based on customer feedback received daily. Nowadays, this feedback data is accessible on social media platforms, significantly aiding in analyzing customer sentiments [2]. Organizations are investing significant resources in research to create tools and strategies for analyzing customer feedback data and aligning products with current market trends [3]. Likewise, banking news reflects the general atmosphere or attitude conveyed in articles, reports, and discussions rooted in economic sentiment. Researchers increasingly favor text-based economic activities [4] due to their advantages over surveys in terms of cost, scope, and timeline.

Constraints imposed and integrity issues associated with text-based sources like news, microblogs, and organizational product disclosures have limited research in this field when analyzing their impacts on various market aspects [5]. Sentiment analysis represents a distinctive feature within the financial sector, enabling sentiment analysis that gauges customer confidence in banking products and services. Banks can pinpoint recurring issues or concerns raised by customers and promptly address them, thus enhancing customer satisfaction and loyalty. Utilizing feedback data allows for enhancing existing products and developing new services tailored to customer preferences while bolstering the organization's reputation and competitive offerings. Identifying unusual patterns and sentiments also serves the purpose of detecting fraud and can be harnessed for customer awareness [6]. Similarly, favorable banking news increases investor confidence, contributes to an upward market trend, and enhances an organization's reputation. Conversely, negative information can have the opposite effect [4]. Public sentiment regarding regulations and policies can sway lawmakers and regulators, shaping the formulation of crisis management strategies.

Contextual comprehension significantly influences sentiments; for instance, 'interest' typically evokes positive sentiment, while 'loans' often yield negative sentiments. Challenges arise in sentiment analysis due to sarcasm, subjectivity, and multilingual nuances in language. Within financial institutions, data often holds sensitive information, making balancing privacy regulations vital when handling customer feedback, reviews, and media releases. Leveraging the integration of machine learning and Natural Language Processing (NLP) [7] techniques proves instrumental in mitigating these challenges.

In this paper, various sources known for their integrity gather the data. After pre-processing and clustering the data, the system computes sentiment scores. The analysis system ensures the integrity of the document's content, focusing on preserving form rather than just content. This approach helps to prevent different classes resulting from the synonyms and antonyms in the text document. The conservation block compares the normalized sentences of the analysis system with newly generated sentences containing synonyms and antonyms, utilizing a control distance mechanism. This process results in an updated dataset for the experiment. The analysis system provides the updated dataset as input to the ELECTRA, a self-supervised language representation learning model for classifying responses within the context of the proposed customer-based banking analysis system.

The rest of the paper includes Section II, which elaborates on banking products and services. Section III presents related work on sentiment analysis, focusing on customer feedback and news from diverse sources. Section IV delves into the proposed system model and its architecture. The research methodology of the model is discussed in Section V, followed by result analysis and discussion of the proposed hypothesis with novelty, strength, and implication in Section VI. Finally, Section VII summarizes the paper, providing conclusions and future scope prospects.

II. Banking Products and Services

The market offers a variety of banking products, each differing from one organization to another. These products and services undergo regular updates and enhancements. The banking system has two primary categories: retail banking and corporate banking. Table I provides an overview of the products and associated services offered by retail and corporate banking sectors.

TABLE I. Products of Retail and Corporate Banking

Retail Banking	Corporate Banking
Checking and saving account, Certificates of deposit, Mortgage, Automobile financing, Credit cards, Lines of credit (Home equity lines of credit and personal credit products), Foreign currency and remittance services, Stock brokerage, Insurance, Wealth management, Private banking.	Loans and other credit products, Treasury and cash management services, Equipment lending, Commercial real estate, Trade financing, Employer service.

III. Related Work

Sentiment analysis assesses customers' responses to products, services, and situations by analyzing texts, posts, reviews, news, and other digital content. This analysis aids business leaders in comprehending customer attitudes and market perceptions over time.

A. Sentiment Analysis on Banking Data

The bank ontology facilitates extracting text feedback data features from websites like Mouthshut.com and myBankTracker.com. The experiment then applies sentiment classification to this data [8]. Rule-based classifier helps to analyze the sentiments of Russian review texts to classify sentiments [9]. The Vader Aware Dictionary and Sentiment Reasoner (VADER) [10] is a lexicon and rule-based sentiment analysis tool specifically crafted to discern sentiments within social media data related to UniCredit bank's European region. Retail banks in South Africa employ both lexicon-based and machine learning-based methods [11] to assess customer feedback sentiments. The results from a fine-tuned DistilBERT model are fed into machine learning classifiers such as Random Forest, Decision Tree, Logistic Regression, and Linear Support Vector Classifier (SVC) [12] to categorize sentiments in news related to Indian banking, governmental, and global topics. Nine classifiers, including Naïve Bayes, Logistic Regression, K-Nearest Neighbours, Support Vector Machines, Random Forest, Decision Tree, Adaptive Boosting, eXtreme Gradient Boosting, and Light Gradient Boosting Mechanism [13], have been employed to detect customer satisfaction levels within Indonesian banks, such as Jenius, Jago, and Blu. The BERTopic architecture utilizes a combination of Kernel Principal Component Analysis (KernelPCA) and K-means Clustering to generate coherent topics, similar to the Latent Dirichlet Allocation (LDA) [14] approach. This method calculates coherence scores for Nigerian bank data, facilitating sentiment analysis. Word representation has evolved by integrating static and contextual words to handle language ambiguity, encompassing semantics, and syntax within a given context. These word representations are fed into a Convolutional Neural Network (CNN) [15] to capture sentiments within financial news contexts.

B. Sentiment Analysis Approach on Text Data

The RCNN [16] analyzes each word's context in the document and then applies the max-pooling layer to identify the text's crucial elements for classification, along with the SoftMax layer. Clustering techniques, such as the K-means-type algorithm [17], exhibit improved performance when applied to balanced data, whereas the designed weighting model delivers exceptional results for both balanced and unbalanced datasets. The word embedding layer [18] converts sentences into words, preserving the contextual information of each word, and then applies a CNN for sentiment analysis. Support Vector Machine (SVM), Deep Learning (DL), and Naïve Bayes (NB) [19] classifiers utilize sentiment scores and the associated weights of hashtags to classify sentiments within social media data. The classification of sentiments [20] involves employing the Bidirectional Long Short-Term Memory (BiLSTM) layer, a global pooling mechanism, and a sigmoid layer. The Collaborative and Bidirectional Gated Recurrent Unit (BiGRU) can also adopt this approach to assess performance. Hierarchical Attention Networks (HAN) [21] enable the identification of sentiment polarity in customer communications, thereby enhancing the efficiency of Customer Relationship Management (CRM) operators in terms of response time. The integration of the BERT model with Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU) algorithms [22] facilitates the analysis of positive, negative, and neutral sentiments.

The literature analysis above demonstrates various modelling approaches employed in sentiment analysis of banking text data, incorporating different types of text data and data corpora. Examining synonyms and antonyms within a data corpus is instrumental in simplifying the identification of customer sentiments toward financial organizations in diverse regions. Further research is necessary to develop a more efficient approach capable of swiftly discerning customers' intent, serving as a driving force to create a more streamlined sentiment analysis context for optimizing banking business strategies.

IV. System Model

The proposed sentiment analysis approach comprises two key components, as illustrated in Fig. 1. The initial phase involves collecting banking data from various sources, including customer service teams, visual feedback tools, review sites, net promoter sources, online surveys, social media, and news media facilitated by banking organizations. These data are then gathered and stored in a database. A "Stop-word" data cloud developed helps to preprocess the dataset documents.

The filtered and pre-processed text data identifies the sentiment scores for unique words. The subsequent step involves determining synonyms and antonyms for these unique words and their corresponding sentiment scores. Replacing each unique word with synonyms and antonyms forms new sentences, as depicted in Fig. 1. This process allows considering a maximum of three synonyms and antonyms for each identified word during replacement.

The experiment employs the Control Distance approach to assess the similarity between the original and newly generated sentences by rephrasing with synonyms and antonyms. Based on their degree of similarity, the experiment uses the resulting similarity score to categorize sentences as positive, negative, or neutral.

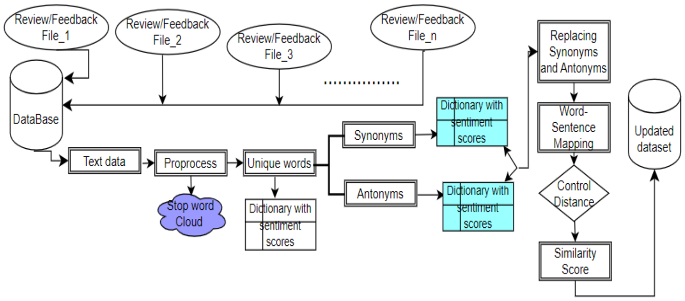


Fig. 1. System model to create an updated dataset.

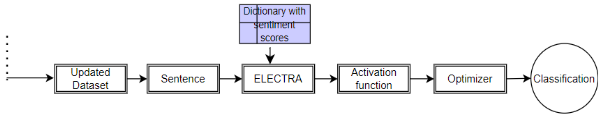


Fig. 2. System model to identify sentiment.

A self-supervised language representation learning model, ELECTRA, trains itself using all words from the data dictionary with sentiment scores. This model is then applied to the updated dataset, as Fig. 2 illustrates. Subsequently, individual sentences are classified and identified for sentiment, enhancing the efficiency of the Sentiment Analysis framework. ELECTRA operates efficiently on sample words, making it faster and requiring fewer resources. The following provides a summary of the detailed implementation:

- The initial step involves gathering datasets from a variety of banking data sources.
- Design and develop the 'Stop-word' word cloud that does not impact sentiments.
- Then, pre-processed the text data using the specially designed 'Stop-word' word cloud.
- Create a data dictionary by identifying unique words and their corresponding sentiment scores.
- The positive, negative, and neutral sentiment scores determine synonyms and antonyms of unique words and establish a distinct data dictionary for these synonyms and antonyms.
- A control distance approach, as proposed, assists in measuring the similarity score between the original and updated sentences. The resulting dataset encompasses positive, negative, and neutral sentiments.
- The ELECTRA model is self-trained and then applied to the updated dataset by merging the data dictionary containing synonyms, antonyms, unique words, and their sentiment scores. SoftMax activation and a modified Adam optimizer help to identify the sentiment for each feedback statement.

V. Research Methodology

This section delves into the analysis of sentiments from banking feedback and media data, encompassing a discussion of the experiments and the results obtained. Firstly, it covers the specifics of the dataset employed for training, cross-validation, and testing, followed by a detailed examination of the approach's implementation. Following this, the analysis delves into the experiment results, revealing the sentiments and information patterns present within the datasets. The study evaluates the effectiveness and performance of the proposed approach by comparing it with other established methods. Furthermore, the analysis includes applying the predefined approach to the post banking customer feedback and phrase banking datasets to determine the accuracy score, given their absence of prior implementation using the ELECTRA model.

A. *Dataset Collection*

Central banks worldwide gather customer feedback on banking services, encompassing satisfaction levels, complaints, and insights on product usability. The sources of post bank customer data [23] are from the Russian finance website "www.banki.ru" spanning 2013 to 2019. Reviews are rated from one to five, with one denoting negativity and five indicating positivity. The "responses_header" column comprises feedback messages from 16,659 customers.

The phase bank dataset [24] includes positive, negative, and neutral sentiments of 5,000 customers about companies listed in OMX Helsinki. The data originates from the LexisNexis database, including 10,000 randomly selected articles from limited financial and economic resources. Analyzing these sentiments assists the marketing team in improvising the banking products and services. The details of collected datasets are shown in Table II and saved in CSV files to calculate the sentiment content of each text sentence. These datasets are divided into 65% for training, 15% for cross-validation, and 20% for testing.

TABLE II. The Datasets in the Experiment

sl#	Dataset	Provided Ratings
1	Post Bank Customer Review	Positive: 2160
2		Negative: 2472
3		Neutral: 1391
4	Phrase Banking	Positive: 1853
5		Negative: 861
6		Neutral: 3131

B. Text Data

Each dataset includes a variety of text messages shared by different banking customers and media sources, extracting subjective information from the shared data. The goal is to discern the attitude, emotional tone, market sentiment, and expressions within a text, allowing for the analysis of overall sentiments from individual feedback and media statements. Various methods, such as rule-based, machine learning, and deep learning techniques, are employed to assess these sentiments.

The article employs a rule-based approach to measure the sentiment scores of unique words in the dataset file. It creates data dictionaries that assign sentiment scores to unique words, synonyms, and antonyms. These dictionaries help to determine the overall sentiment scores of sentences or paragraphs.

C. Pre-processing the Text Data

Text pre-processing involves cleaning text data by eliminating irrelevant information like URLs, numbers, and punctuation marks. This step ensures the availability of more pertinent texts for conducting sentiment analysis activities. Then, the process [25] involves tokenizing, normalizing, removing stop words, stemming, and lemmatization.

In this implementation, aside from general stop words, specific banking-related stop words such as currency names (Rupee, dollar, etc.), Roman numerals (I, II, III, etc.), and auditing firms (KPMG, Deloitte, etc.) are utilized. These stop words significantly impact sentiment determination and play a significant role in determining sentiment. They are used to generate the word cloud, which, in the process, assists in removing these words from the dataset files, as shown in Fig. 3. In pre-processing and stop word removal, the process extracts unique words, facilitating the identification of context and reducing words to their root form. Identifying distinct words in a dataset facilitates the analysis of individual words for information retrieval purposes. The extraction of unique words serves several functions, including:

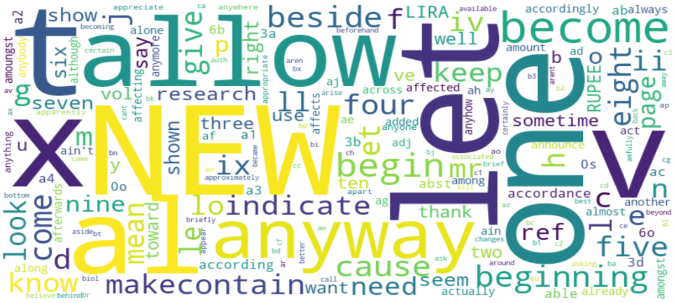


Fig. 3. Generated word cloud for stop words.

- Analyzing the richness of vocabulary in the data corpus.
- Constructing a data dictionary.
- Accessing the diversity and complexity in the written content.

WordNet, a computational lexical database with linguistic and physiological features [26], identifies English verbs, nouns, adjectives, and adverbs through part-of-speech tagging. Table III illustrates some unique words extracted from the data corpus.

TABLE III. Few Unique Words from the Data Corpus

solutions	powerful	multimedia	company	relevant
sales	search	technology	moved	doubled
turnover	platform	location	leverage	strengthened
content	communities	domestic	profit	loss
period	project	recorded	ownership	generated

D. Sentiment Score and Data Dictionary

Determining the significance of specific words within a group of relevant words is vital in determining the sentiment scores. The newly proposed approach calculates this importance by assigning sentiment scores, following the formulation, as depicted in Eq. (1).

$$SC = WP * \log(n/N) \tag{1}$$

where, SC is the sentiment score, WP is the word presence, n is the frequency of the unique word W_i, and N is the number of words present in the document. As $\log()$ is used in the formulation, SC value will be less and vice versa if the frequency of the word is more. WP takes the value as 0 or 1 for the presence or absence of the words in the document.

It can be challenging to obtain unique words and their corresponding sentiment scores for the next step. Therefore, a metadata repository, termed a data dictionary, is established to store words and their associated sentiment scores (SC). This data dictionary elucidates each word's context, offering crucial information for future reference without the need for in-depth analysis of the raw data. The SC value categorizes words as positive, negative, or neutral based on predefined threshold values, denoted as L_{value} and H_{value} in Eq. (2).

$$\begin{cases} positive & \text{if } SC \leq L_{value} \\ negative & \text{if } L_{value} < SC > H_{value} \\ neutral & \text{if } SC \geq H_{value} \end{cases} \tag{2}$$

The final step involves clustering the banking data document's positive, negative, and neutral words to serve as input for the subsequent phase.

In this implementation, words with a threshold value below and equal to 0.003 are classified as "positive," while words between 0.003 and 0.007 are considered "negative," and words exceeding the threshold of 0.007 are "neutral." These values are pivotal in creating a new dataset sorting statements into positive, negative, and neutral categories and establishing a data dictionary for positive, negative, and neutral categories containing unique words and their corresponding sentiment scores. Table IV presents a selection of unique words with their sentiment scores.

TABLE IV. UNIQUE WORDS WITH SENTIMENT SCORES

Sl#	Unique words	Sentiment Score
1	profit	0.0021
2	search	0.0014
3	relevant	0.0043
4	multimedia	0.0055
5	platform	0.0081

E. Synonyms and Antonyms with Sentiment Scores

Synonyms are words or phrases sharing identical meanings, allowing their replacement in a specific context without altering the meaning. Antonyms represent words with opposite meanings, conveying contrasting ideas in specific contexts. Comprehending synonyms and antonyms is vital for language and communication, fostering a diverse vocabulary and enhancing effective expression. This knowledge equips writers and speakers to convey messages precisely, highlighting contrasts and differences between ideas and concepts. WordNet [27] discerns connections between word meanings, identifying relationships such as synonyms and antonyms. For instance, “move, drive, impel” are synonymous words, while “stay, stop, discourage” are antonyms.

In this implementation, the process extracts synonyms and antonyms from unique words. It then allocates positive sentiment scores to synonyms, retains the same scores for unique words, and assigns antonyms a negative sentiment score with an identical value. Table V presents a selection of synonyms and antonyms with their sentiment scores corresponding to unique words.

TABLE V. SYNONYMS AND ANTONYMS WORDS WITH SENTIMENT SCORES

Sl#	Unique words	Synonyms	SC	Antonyms	SC
1	profit	financial gain	0.0021	loss	-0.0021
2	search	explore	0.0014	ignore	-0.0014
3	relevant	pertinent	0.0043	irrelevant	-0.0043
4	multimedia	multimodal	0.0055	monomodal	-0.0055
5	platform	podium	0.0081	ground	-0.0081

F. Sentence Similarity and Updated Dataset

Substituting synonyms and antonyms results in an updated dataset derived from the existing one. The newly proposed method, Control Distance, calculates sentence similarity after rephrasing these words, as depicted in Fig. 4.

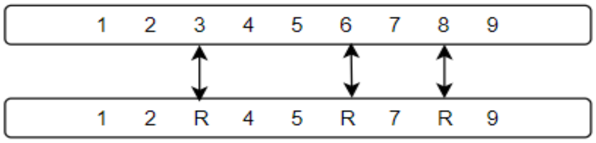


Fig. 4. Replaced words on original statement.

1) Control Distance: Calculating the sentence similarity between two strings involves determining the minimum number of edits (insertions, deletions, or substitutions) necessary to transform one string into another. This method also handles topographical inconsistencies in string data. In this implementation, the proposed expression assesses the similarity or dissimilarity between the original sentence and the sentences with substituted synonyms and antonyms.

$$SS_v = n * \frac{\sum_{i=1}^{n} SC_i}{N} \tag{3}$$

Here, SS_v is the Sentence similarity, and n is the words substituted in a sentence or a paragraph. N is the sum of the words present in a sentence or a paragraph. SC_i is the sentiment score of the substituted words.

After determining the Sentence Similarity (SS_v) between the original sentence and the sentences substituted with synonyms and antonyms, the SS_v is utilized to classify the sentences as similar or dissimilar by applying a standard threshold value, T_v.

$$Similarity = \begin{cases} Similar & SS_v \geq T_v \\ Dissimilar & SS_v < T_v \end{cases} \tag{4}$$

The similarity of the sentences determines the corresponding positive, negative, and neutral sentiments assigned to the original sentences.

Standard threshold values help measure sentences’ positive, negative, and neutral sentiments. Sentences with similarity values ranging from 0 to 3.0 are classified as positive, while those with similarity scores between 0 and -3.0 are negative. The remaining sentences with different similarity scores are considered neutral. Table VI presents a few of the sentence similarity scores.

TABLE VI. SIMILAR AND DISSIMILAR SENTENCES

Sl#	Statement	Similarity Score	Sentiment
1	L&G still paying price for dividend cut during crisis, chief says	Nill	Positive
2	L&G still rewarding price for dividend cut during emergency, chief says	2.26	Positive
3	L&G still remunerating price for dividend cut during disaster, chief says	1.93	Positive
4	L&G still compensating price for dividend cut during hardship, chief says	2.67	Positive
5	L&G still forgiving price for dividend cut during normalcy, chief says	-1.63	Negative
6	L&G still withholding price for dividend cut during certainty, chief says	-0.98	Negative
7	L&G still deferring price for dividend cut during harmony, chief says	-2.16	Negative

G. Model Implementation with Optimizer

ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately) [28] stands as a self-supervised language representation learning model. Its pretraining objective resembles the traditional Masked Language Model (MLM) but incorporates binary classification objectives. ELECTRA is a binary classifier during pretraining but can help adapt for multiclass classification tasks through finetuning.

ELECTRA operates through a Discriminator (D) designed to differentiate between “real” and “forged” tokens within a sentence. Additionally, it includes a Generator (G) network that substitutes certain input tokens with incorrect ones. Both the generators and discriminators are composed of transformer encoder layers, following the formulation below.

- Masked token replacement with sentiment score: The sequence token $T = [t_1, t_2, t_3, t_4, ..., t_i, ..., t_n]$ with associated sentiment score $SC = [SC_1, SC_2, SC_3, SC_4, ..., SC_i, ..., SC_n]$ where t_i represents the i^{th} token with corresponding sentiment score SC_i. The process of generating masked token operates as follows: for token t_i, it is replaced by a mask token $[MASK]$ with sentiment score SC_i.

- Generator (G): The generator (G) attempts to predict original token t_i from the $[MASKED]$ token, and its output is a probabilistic distribution over the vocabulary of each position i. Generator Loss (L_G) is calculated as

$$L_G = \sum_{i=1}^{n} SC_i * \sum_{j=1}^{V} t_i \log G([MASK], t_{connect}) \quad (5)$$

Here, V represents the vocabulary size, $G([MASK], t_{connect})$ signifies the output probabilities of the generator for masked taken $[MASK]$ given the context $t_{connect}$.

$$L_D = \sum_{i=1}^{n} SC_i * t_i \log D(t_i, t_{connect}) + (1 - t_i) \log(1 - D([MASK], t_{connect}))) \quad (6)$$

Here, $D(t_i, t_{connect})$ represents discriminator prediction for the token t_i with given context $t_{connect}$. Both the generator and discriminator are trained to minimize their respective loss based on the Sentiment scores.

To produce a probabilistic distribution over masked token, the SoftMax Function [29] is employed in the generator component. The function converts raw score denoted as $[S_1, S_2, S_3, S_4, ..., S_i, ..., S_v]$ into probabilities $[P_1, P_2, P_3, P_4, ..., P_i, ..., P_v]$ using below formula

$$SoftMax(P_i) = \frac{e^{S_v}}{\sum_{j=1}^{V} e^{S_j}} \quad (7)$$

The raw score generated by the ELECTRA generator is passed through the SoftMax function to obtain the probability of each token being the correct replacement for the masked token.

An optimizer is a crucial algorithm employed to fine-tune neural network attributes, such as learning rates and weights, to minimize loss during training. Its main purpose is to identify the optimal parameters that minimize the disparity between predicted and actual values. In this particular implementation, using the Adaptive Moment Estimation (ADAM) optimizer [30] proves invaluable for obtaining optimal results from the ELECTRA model. Additionally, integrating the L2 regularizer [31] into the modified Adam optimizer enhances convergence, addresses sparse gradient issues, prevents computation in local minima, and handles imbalanced data effectively.

H. Sentiments

Banking customers convey feelings about specific media data, brands, products, or services, encompassing attitudes, opinions, and emotions. Analysts examine sentiments based on written or voice feedback provided by customers and the mass media data, typically classifying them into three categories: positive, negative, and neutral [17], reflecting different emotional tone.

- **Positive:** Expresses satisfaction, joy, love, excitement, etc.

- **Negative:** Conveys feelings of sadness, anger, disappointment, displeasure, etc.

- **Neutral:** Lacks strong positive or negative emotions, presenting factual text without any specific emotional tone.

An automated sentiment analysis approach that identifies these emotions in feedback and media text documents offers valuable insights into the emotional content, proving particularly useful in banking data analysis.

I. Algorithm for Sentiment Analysis

The suggested approach for analyzing sentiment from textual data involves four distinct phases. Algorithm 1 addresses text data pre-processing. Algorithm 2 outlines the identification of Synonyms and Antonyms associated with individual words and their respective Sentiment Scores. Algorithm 3 illustrates the generation of an enhanced Dataset. Lastly, Algorithm 4 demonstrates sentiment classification utilizing the ELECTRA model on the updated dataset.

Algorithm 1 Pre-Processing the text data

Notations: Banking_Feedback_Corpus (Tokens, Sentences, Sentiment, Threshold_1, Thrreshold_2)
INPUT: TextFile (Banking_Feedback_Corpus_File)
OUTPUT: Sentiment= (Positive, Negative, Neutral)

```
Begin
Read text data from Banking_Feedback_Corpus_File
Remove the URLs, Numbers, Punctuations
Words ← Split the Banking_Feedback_Corpus_File by Space
words ← lower (Words)
Add “words” to “processing_words” library
Create “Stop words” Word Cloud
Find distinct stemming words to the “stemmed_word” container
Add “stemmed_word” to “processing_text” library
```

VI. Results and discussion

The Sentiment Analysis model uses Python, relying on various frameworks and machine learning libraries for implementation. Python (specifically Python 3.6.3$rc1$) and the NLTK 3.0 library are used for text data processing, providing a range of built-in functions. The Seaborn and Matplotlib libraries for statistical data analysis and visualization help plot the graph. Additionally, the experiment utilizes machine

Algorithm 2 Extracting Synonyms and Antonyms from Unique words with Sentiment Score

```
unique_words ← set (processing_text)
unique_words_list ← list (unique_words)
Read the “processing_words” library
for word in unique_words_list do
    X ← n/N
    Y ← log (X)
    SC ← WP * Y
    if $SC \geq Threshold\_1$ then
        Positive Sentiment
    else if $SC \geq Threshold\_1$ AND $SC \leq Threshold\_2$ then:
        Negative Sentiment
    else
        Neutral Sentiment
    end if
end for
Create data Dictionary words with Sentiment Scores
synonyms, antonyms ← get_synonyms_antonyms (unique_words_list)
Create data dictionary Unique word, Synonyms, and Positive SC
Create data dictionary Unique word, Antonyms, and Negative SC
```

Algorithm 3 Creating Updated Dataset

```
Replace three synonyms and antonyms words on original sentence from text file
Original Sentence
sentence_to_compare= [sen_1, sen_2, sen_3, sen_4, sen_5, sen_6]
for sentence in sentence_to_compare do
    n ← number of replaced words in Original Sentence
    SS= []
    for i= 1 to n do
        Read $SC_i$
        Sum= Add ($SC_i$)
        SS = n * (Sum/N)
        if $SS > Similarity\_value$ then
            Similar
        else
            Dissimilar
        end if
    end for
end for
Update the sentiment of the sentences with similarity score
```

Algorithm 4 Sentiment Analysis using ELECTRA

```
Replace three synonyms and antonyms words on original sentence from text file
Segregate the dataset to train, cross-validation, and test
Self-Train the model with words, and the sentiment scores
ELECTRA Modelling applied Sentence/Paragraph
OUTPUT: Sentiment= Positive, Negative, Neutral
Measure Accuracy
End
```

learning libraries such as Keras (version 2.6.0), TensorFlow (version 2.6.0), and Scikit-learn.

A. Results

In this setup, the assessment of the model's validation performance uses various metrics from the confusion matrix, including accuracy, recall, precision, and F-score. A rigorous $10 - fold$ cross-validation approach helps to evaluate the sentiment scores. Multiple scenarios, including the original dataset, datasets with synonyms replaced, datasets with antonyms replaced, and combinations help to estimate the model's performance. As a constraint, the implementation includes only three synonyms and antonyms words. Table VII displays the model performance on the Postbank dataset, as well as datasets containing synonyms and antonyms rephrased sentences.

Fig. 5 illustrates a graphical representation of negative, neutral, and positive sentiments. The basis of this representation is on performance measures such as False Positive Rate (FPR) and False Negative Rate (FNR) values. The analysis includes the post-bank dataset, datasets with replaced synonyms and antonyms. The analysis of the post-bank dataset involves a combination of sentences with rephrased synonyms and antonyms, along with the original sentences. Table VIII presents the performance metrics for the post-bank dataset, including sentences with rephrased synonyms and antonyms.

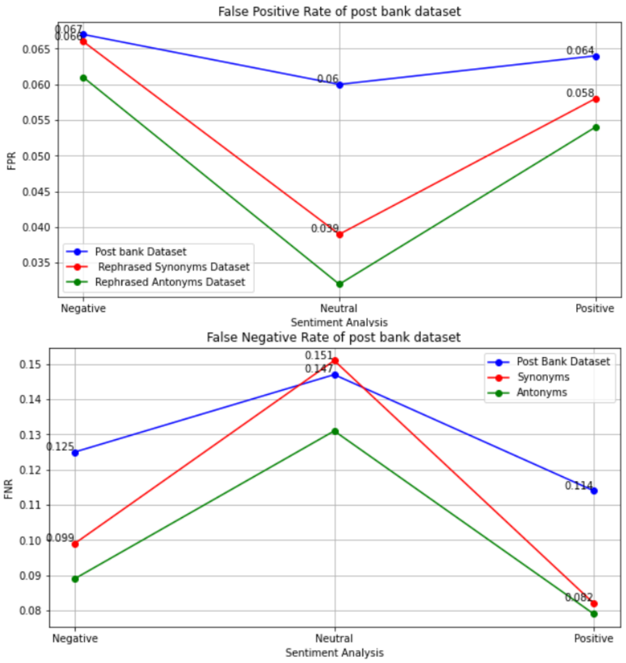


Fig. 5. FPR and FNR values for post bank, rephrase synonyms and antonyms dataset.

Regarding the performance metrics, neutral sentiments are exhibiting lower performance compared to other sentiments. Similarly, the assessment of the performance of the Phrase bank dataset includes sentences with rephrased synonyms and antonyms. Table IX displays the model performance on the Phrase bank dataset, as well as datasets containing synonyms and antonyms rephrased sentences.

TABLE VII. PERFORMANCE METRICS OF POST BANK, SYNONYMS AND ANTONYMS DATASETS

Sl#	Parameters	Original Dataset			Synonym Dataset			Antonym Dataset		
		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
1	Precision	87.26	84.80	88.98	87.16	90.41	89.81	88.23	92.01	90.50
2	Recall	87.40	85.22	88.50	90.05	84.83	91.72	91.07	86.81	92.07
3	F1 score	87.33	85.01	90.76	88.74	87.53	90.76	89.63	89.34	90.28
4	Support	1176	1068	1269	1176	1068	1269	1177	1062	1274

TABLE VIII. PERFORMANCE MEASURE OF POST BANK CUSTOMER DATA ON TESTING DATA

Pİndex	Negative	Neutral	Positive	Accuracy	MȦvg	WȦvg
Precision	89.8765	95.6587	91.5895	92.1577	92.3749	0.922754
Recall	93.1741	88.3003	94.5440	92.1577	92.0061	0.921577
f1 Score	91.4956	91.8323	93.0433	92.1577	92.1237	0.921528
Support	2344	2171	2511	92.1577	7026	7026

Fig. 6 depicts a graphical representation of negative, neutral, and positive sentiments based on performance metrics such as False Positive Rate (FPR) and False Negative Rate (FNR) values. The analysis of the Phrase bank dataset involves a combination of sentences with rephrased synonyms and antonyms, along with the original sentences. Table X presents the performance metrics for the Phrase bank dataset. The proposed model analyzes both the Post bank customer data and Phrase bank datasets. These datasets contain statements with rephrased synonyms and antonyms alongside original statements or paragraphs. The analysis utilizes the modified ADAM optimizer. Fig. 7 illustrates the overall performance of the proposed model for both datasets.

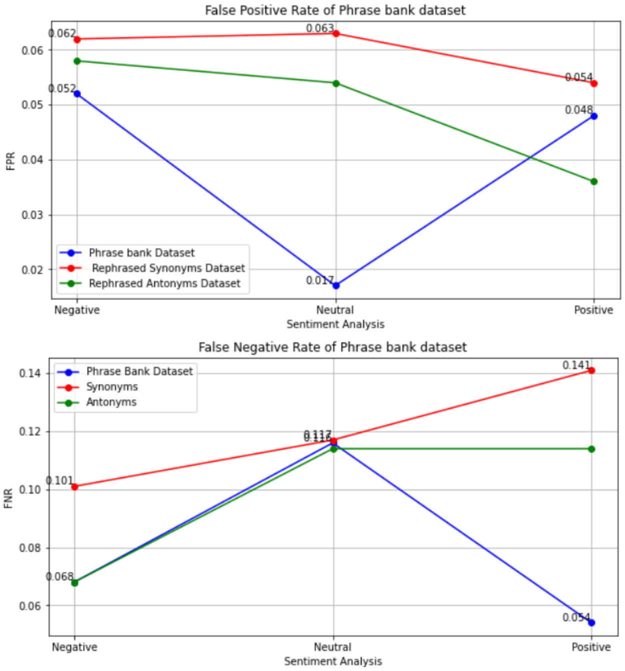


Fig. 6. FPR and FNR values for phrase bank, rephrase synonyms and antonyms dataset on testing data.

The implementation is validated using the ADAM optimizer and the modified Adam optimizer, incorporating L2 regularization. The modified Adam optimizer outperforms the standard Adam optimizer significantly. Table XI shows the comparative performance results.

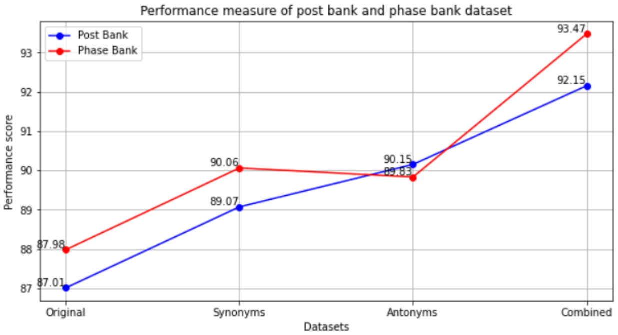


Fig. 7. Overall performance of the post bank and phrase bank dataset on testing data.

B. Performance Comparison

Utilizing the banking dataset in the proposed model yields superior results to conventional datasets such as general English sources (Wikipedia and BooksCorpus) and banking datasets like Phrase Bank.

The Name Entity Recognition (NER) task employs the financial language model based on ELECTRA [32]. In NER, a knowledge graph assists in comprehending the connections among various financial entities, such as individuals, organizations, and locations. The FLANG model achieves 82% accuracy in NER when applied to general English datasets like Wikipedia and BooksCorpus. However, the NER and FLANG concepts are also applied to the post-bank dataset, resulting in an accuracy of 82.9%.

The proposed approach is applied to analyze sentiments in the financial Phrase Bank dataset and financial tweets [33] utilizing the FinBERT model. Incorporating the CLS token, a special marker used for sequencing classification task representations, the model achieves an accuracy of 87.1%. Implementing the Phrase bank dataset [24] involves applying the ELECTRA model with Self-Attention and prediction layers, resulting in an accuracy of 83.87%, as shown in Table XII. This experiment represents the initial utilization of the Phrase bank dataset with ELECTRA and CLS tokens. Before this, ELECTRA had not incorporated the Phrase bank dataset.

C. Strength and Implication

The implementation results have numerous practical and theoretical implications. Theoretically, this research activity enhances our understanding of the sentiments conveyed by individual words in a sentence. It sheds light on market

TABLE IX. Performance Metrics of Phrase Bank, Synonyms and Antonyms Datasets

Sl#	Parameters	Original Dataset			Synonym Dataset			Antonym Dataset		
		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
1	Precision	87.75	87.75	88.47	88.88	89.34	92.13	87.68	90.97	90.50
2	Recall	89.85	88.26	85.80	93.13	88.56	88.51	90.64	88.66	90.23
3	F1 score	88.79	88.01	87.11	90.96	88.95	90.29	89.13	89.80	90.59
4	Support	335	341	331	670	682	662	1005	1023	993

TABLE X. Performance Measure of Phrase Bank Data on Testing Data

Pİndex	Negative	Neutral	Positive	Accuracy	MȦvg	WȦvg
Precision	93.7876	93.2171	93.4410	93.479	93.4819	93.4805
Recall	93.1343	94.0371	93.2528	93.479	93.4747	93.4790
f1 Score	93.4598	93.6253	93.3468	93.479	93.4773	93.4787
Support	1005	1023	993	93.3468	3021	3021

TABLE XI. Comparison of Model Performance on ADAM and Modified ADAM Optimizer

Sl#	Dataset	Parameters	ADAM	Modified ADAM
1	Post Bank Customer Review	Original Statements	81.9	87.01
2		Statements having Only Synonyms	83.19	89.07
3		Statements having Only Antonyms	82.45	90.15
4		Original + Rephrased statements	87.15	92.15
5	Phrase Banking	Original Statements	82.79	87.98
6		Statements having Only Synonyms	83.54	90.74
7		Statements having Only Antonyms	86.17	89.83
8		Original + Rephrased statements	87.36	93.47

TABLE XII. Performance Comparison of the Proposed Model

Sl#	Model Implementation	Accuracy
1	FLANG˙ELECTRA (NER) [32]	82.0%
2	FLANG_ELECTRA (Post Bank customer dataset, NER)	82.9%
3	FinBERT (Phrase bank, CLS token) [33]	87.1%
4	ELECTRA (Phrase bank, CLS token)	83.87%
5	Proposed Model (Post Bank Customer Dataset)	92.27%
6	Proposed model (Phrase bank dataset)	93.48%

sentiments and aids in comprehending the spread of misinformation within the Banking domain. This concept offers valuable insights into the psychological and social factors that influence the dissemination and reception of misinformation, informing the design of marketing strategies.

In practical terms, the suggested approach for evaluating sentiment scores assists in discerning the sentiments of individual meaningful words. Using synonyms and antonyms enhances the system's efficiency, enabling it to identify sentiments across a broader spectrum of arguments. Additionally, the control distance approach aids in recognizing similarities between sentences and their rephrased counterparts.

D. Novelty and Scope

The study centers on extracting sentiments from banking text data, including feedback and media reports. Previous literature has covered various studies on sentiment analysis. This research is purpose-built for improved efficiency and reduced latency. The paper's novelty can be summarized as follows:

- Proposed a sentiment-scoring approach for individual words, synonyms, and antonyms. Assign sentiment scores (positive, negative, and neutral) to synonyms and antonyms.
- Devised a Control Distance approach to validate the similarity score between the original sentence and the synonyms and antonyms used in sentences.
- Subsequently, design a self-trained model for the ELECTRA model, incorporating words with sentiment scores from the designed data dictionary.
- The implementation assessed performance using the ADAM optimizer and the modified ADAM optimizer, incorporating the L2 regularizer.

VII. Conclusion and Future Scope

The study presents a comprehensive sentiment analysis, identifying positive, negative, and neutral sentiments within customer feedback, financial, and economic texts. This analysis aids in combating misinformation in the market and informs marketing strategies. The article employed several methodologies, including text data preprocessing, sentiment identification for each word, synthesizing synonyms and antonyms for unique words, labeling individual sentences or paragraphs, and modeling. The dataset also labeled using sentiment scores for individual words and the control distance technique, providing meaningful sentiment analysis for individual sentences, thereby enhancing the overall implementation.

The implementation aims for a profound understanding of sentiments across a diverse dataset. A normalized threshold range of sentiment scores facilitates categorizing unique words into positive, negative, and neutral labels. Synonyms and antonyms are assigned similar sentiment scores, distinguished by positive and negative signs, respectively. The control distance approach verifies the sentiment scores of individual words in a sentence, evaluating positive, negative, and neutral statements. Subsequently, the ELECTRA model is self-trained using words with sentiment scores to classify sentiments. The implementation evaluates the model's output using both Adam and modified Adam optimizers for comparison.

The implemented approach demonstrates exceptional performance compared to FLANG_ELECTRA with NER model and the FinBERT model, using both Post banking customer data and Phrase banking data. The performance improvement compared to traditional models is substantial.

This article summarizes exploring the sentiments expressed in banking and financial-related data from customer feedback and financial news. The findings provide valuable insights for policymakers in the banking and financial sectors, aiding in developing strategies for customer feedback analysis, brand monitoring, product and service evaluation, fraud detection, market research, and competitor analysis. Additionally, the research holds practical significance by informing the design of automated tools for identifying sentiments within banking and financial text data.

Implementation of the proposed approach focuses on capturing sentiment scores within the documents. However, future work can broaden the scope. Currently, this implementation uses only three synonyms and antonyms. Exploring more synonyms and antonyms for word replacement opens avenues for further research.

Moreover, incorporating diverse deep learning, machine learning, and other models can enhance sentiment accuracy. Despite these options, this approach is the foundation for future investigations involving banking datasets. Researchers can also explore different banking-related datasets for in-depth analysis.

REFERENCES

[1] Mittal, Divya and Agrawal, Shiv Ratan, "Determining banking service attributes from online reviews: text mining and sentiment analysis," "International Journal of Bank Marketing," vol. 40, no. 3, pp. 558–577, 2022.

[2] Gavval, Rohit and Ravi, Vadlamani and Harshal, Kalavala Revanth and Gangwar, Akhilesh and Ravi, Kumar, "CUDASelfOrganizing feature map based visual sentiment analysis of bank customer complaints for Analytical CRM," "arXiv preprint arXiv:1905.09598," 2019, https://doi.org/10.48550/arXiv.1905.09598.

[3] Ahmed, Alim Al Ayub and Agarwal, Sugandha and Kurniawan, IMade Gede Ariestova and Anantadjaya, Samuel PD and Krishnan, Chitra, Akhilesh and Ravi, Kumar, "Business boosting through sentiment analysis using Artificial Intelligence approach," "International Journal of System Assurance Engineering and Management," vol. 13, no. Suppl 1, pp. 699–709, 2022, https://doi.org/10.1007/s13198-021-01594-x.

[4] Shapiro, Adam Hale and Sudhof, Moritz and Wilson, Daniel J, "Measuring news sentiment," "Journal of econometrics," vol. 228, no. 2, pp. 221–243, 2022, https://doi.org/10.1016/j.jeconom.2020.07.053.

[5] Mattia Atzeni, Amna Dridi, and Diego Reforgiato Recupero, "Fine-Grained Sentiment Analysis on Financial Microblogs and News Headlines," Semantic Web Challenges: 4th SemWebEval Challenge at ESWC 2017, Portoroz, Slovenia, May 28-June 1, 2017, pp. 124–128, doi: https://doi.org/10.1007/978-3-319-69146-6_11.

[6] Gandhi, Ankita and Adhvaryu, Kinjal and Poria, Soujanya and Cambria, Erik and Hussain, Amir, "Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions," "Information Fusion," vol. 91, pp. 424–444, 2023, https://doi.org/10.1016/j.inffus.2022.09.025.

[7] Muhammad Taimoor Khan, Muhammad Taimoor Khan, Mehr Durrani, Armughan Ali, Irum Inayat, Shehzad Khalid and Kamran Habib Khan, "Sentiment analysis and the complex natural language," Complex Adaptive Systems Modeling, vol. 4, no. 1, pp. 1–19, 2016, doi: https://doi.org/10.1186/s40294-016-0016-9.

[8] Deepshikha Chaturvedi and Shalu Chopra, "Customers Sentiment on Banks," International Journal of Computer Applications, vol. 98, no. 13, 2014.

[9] Yuliya Bidulya and Elena Brunova, "Sentiment Analysis for Bank Service Quality: a Rule-based Classifier," 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), 2016, pp. 1–4, doi: 10.1109/ICAICT.2016.7991688.fr.

[10] Botchway, Raphael Kwaku and Jibril, Abdul Bashiru and Kwarteng, Michael Adu and Cho vancova, Miloslava and Oplatková, Zuzana Komínková, "A review of social media posts from UniCredit bank in Europe: a sentiment analysis approach," Proceedings of the 3rd international conference on business and information Management, pp. 74–79, 2019.

[11] Kazmaier, J and Van Vuuren, JH "Sentiment analysis of unstructured customer feedback for a retail bank," ORiON, vol. 36, no. 1, pp. 35–71, 2020.

[12] Varun Dogra, Aman Singh, Sahil Verma, Kavita N. Z. and M.N. Talib, "Analyzing DistilBERT for Sentiment Classification of Banking Financial News," Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021, pp. 501–510, 2021, doi: https://doi.org/10.1007/978-981-16-3153-5_53.

[13] Bramanthyo Andrian, Tiarma Simanungkalit, Indra Budi and Alfan Farizki Wicaksono, "Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia," International Journal of Advanced Computer Science and Applications, vol. 13, no. 3, 2022.

[14] Ogunleye, Bayode and Maswera, Tonderai and Hirsch, Laurence and Gaudoin, Jotham and Brunsdon, Teresa, "Comparison of topic modelling approaches in the banking context," Applied Sciences, vol. 13, no. 2, pp. 797, 2023.

[15] Surabhi, Adhikari, Surendrabikram Thapa, Usman Naseem, Hai Ya Lu, Nana Bharathy and Mukesh Prasad, "Explainable hybrid word representations for sentiment analysis of financial news," Neural Networks, vol. 164, pp. 115–123, 2023, doi: https://doi.org/10.1016/j.neunet.2023.04.011.

[16] Siwei Lai, Liheng Xu, Kang Liu and Jun Zhao, "Recurrent Convolutional Neural Networks for Text Classification," Proceedings of the AAAI conference on artificial intelligence, vol. 29, no. 1, 2015.

[17] Baojun Ma, Hua Yuan and Ye Wu, "Exploring performance of clustering methods on document sentiment analysis," Journal of Information Science, vol. 43, no. 1, pp 54–74, 2017, doi: https://doi.org/10.1177/0165551515617374.

[18] Mohammed Attia, Younes Samih, Ali Elkahky and Laura Kallmeyer†, "Multilingual Multi-class Sentiment Classification Using Convolutional Neural Networks," Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018.

[19] Saurav Pradha, Malka N. Halgamuge and Nguyen Tran Quoc Vinh, "Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data," 2019 11th international conference on knowledge and systems engineering (KSE), pp. 1–8, 2019, doi: 10.1109/KSE.2019.8919368.

[20] Zabit Hameed and Begonya Garcia-Zapirain, "Sentiment Classification Using a Single-Layered BiLSTM Model," Ieee Access, vol. 8, pp. 73992–74001, 2020, doi: 10.1109/ACCESS.2020.2988550.

[21] Nicola Capuano, Lica Greco, Pierluigi Ritrovato and Mario Vento, "Sentiment analysis for customer relationship management: an incremental learning approach," Applied Intelligence, vol. 51, pp. 3339–3352, 2021, doi: https://doi.org/10.1007/s10489-020-01984-x.

[22] Amira Samy Talaat, "Sentiment analysis classification system using hybrid BERT models," Journal of Big Data, vol. 10, no. 1, pp. 1–18, 2023, doi: https://doi.org/10.1186/s40537-023-00781-w.

[23] Andrei Plotnikov, Alexey Shcheludyakov, Vadim Cherdantsev, Alexey Bochkarev and Igor Zagoruiko, "Data on post bank customer reviews from web," Data in Brief, vol. 32, pp. 106152, 2020, doi: https://doi.org/10.1016/j.dib.2020.106152.

[24] Malo, Pekka and Sinha, Ankur and Korhonen, Pekka and Wallenius, Jyrki and Takala, Pyry, "Good debt or bad debt: Detecting semantic orientations in economic texts," Journal of the Association for Information Science and Technology, vol. 65, no. 4, pp. 782–796, 2014, doi: https://doi.org/10.1002/asi.23062.

[25] Javed, Muhammad and Kamal, Shahid, "Normalization of unstructured and informal text in sentiment analysis," "International Journal of Advanced Computer Science and Applications," vol. 9, no. 10, 2018, http://dx.doi.org/10.14569/IJACSA.2018.091011.

[26] Atoum, Issa and Otoom, Ahmed, "Efficient hybrid semantic text similarity using WordNet and a corpus," "International Journal of Advanced Computer Science and Applications," vol. 7, no. 9, 2016, http://dx.doi.org/10.14569/IJACSA.2016.070917.

[27] Mukhiya, Suresh Kumar and Ahmed, Usman and Rabbi, Fazle and Pun, Ka I and Lamo, Yngve, "Adaptation of IDPT system based on patient-authored text data using NLP.," 2020 IEEE 33rd international symposium on computer-based medical systems (CBMS), pp. 226–232, 2020, doi: 10.1109/CBMS49503.2020.00050.

[28] Clark, Kevin and Luong, Minh-Thang and Le, Quoc V and Manning, Christopher D, "Electra: Pre-training text encoders as discriminators rather than generators," arXiv preprint arXiv:2003.10555, 2020, doi: https://doi.org/10.48550/arXiv.2003.10555.

[29] Hasan, Mahmud and Islam, Labiba and Jahan, Ismat and Meem, Sabrina Mannan and Rahman, Rashedur M, "Natural Language Processing and Sentiment Analysis on Bangla Social Media Comments on Russia–Ukraine War Using Transformers," Vietnam Journal of Computer Science, pp. 1–28, 2023, doi: https://doi.org/10.1142/S2196888823500021.

[30] Kingma, Diederik P and Ba, Jimmy, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014, doi: https://doi.org/10.48550/arXiv.1412.6980.

[31] Loshchilov, Ilya and Hutter, Frank, "Fixing weight decay regularization in adam," 2018.

[32] Shah, Raj Sanjay and Chawla, Kunal and Eidnani, Dheeraj and Shah, Agam and Du, Wendi and Chava, Sudheer and Raman, Natraj and Smiley, Charese and Chen, Jiaao and Yang, Diyi, "When flue meets flang: Benchmarks and large pre-trained language model for financial domain," "arXiv preprint arXiv:2211.00083" 2022.

[33] Leonardo Colacicchi, Comparison and fine- tuning of methods for Financial Sentiment Analysis. Department of Data Science and Artificial Intelligence, Maastricht University, 2022.

Iterative Learning Control for High Relative Degree Discrete-Time Systems with Random Initial Shifts

Dongjie Chen, Tiantian Lu, Zhenjie Yin*
Basic Courses Department, Zhejiang Police College, Hangzhou, Zhejiang 310053, China

***Abstract*—In this paper, an iterative learning control (ILC) strategy under compression mapping framework is presented for high relative degree discrete-time systems with random initial shifts. Firstly, utilizing the high relative degree of the system and difference term, a control law is designed and a p-order non-homogeneous linear difference equation is established. The appropriate control gain is selected according to the characteristics of solution of the difference equation and the initial shifts, so as to ensure that the high relative degree discrete-time system can reach a steady-state deviation output at a fixed time. Subsequently, a PD-type control law is employed to correct the fixed deviation of the system. Theoretical analysis indicates that this ILC strategy can ensure that the high relative degree systems achieve accurate tracking after the predefined time. Finally, the simulation experiments are conducted on a linear discrete-time Multiple-Input Multiple-Output(MIMO) system with relative degree 1 and a Multiple-Input Single-Output(MISO) system with relative degree 2, respectively, and the results verify the effectiveness of the algorithm.**

***Keywords*—Relative degree; iterative learning control; random initial shifts; difference equation; discrete-time system**

I. Introduction

For the control systems in the field of repetitive operations, iterative learning control (ILC) is a commonly used intelligent control strategy. Drawing from prior batch tracking errors and inputs to update the current batch's control inputs, ILC enables the repetitively operated system to follow the desired trajectory to a high degree of precision over a finite interval. Notably, ILC's lack of reliance on the knowledge of system dynamics, coupled with its superior adaptability, renders it an ideal fit for complex control systems. ILC is widely used in robot control systems [1]-[3], medical rehabilitation [4], [5], multi-agent formation [6], [7], batch processes [8], [9], train automatic control [10], [11]and so on.

The relative degree is utilized to express the extent to which a system control input directly feeds back the system output. In mathematical terms, the relative degree is defined as the lowest order derivative of system output with respect to time, which can be directly fed back by the control inputs. In the discrete-time dynamic systems, the relative degree is manifested as the time delay between the input and output of system, which is inherent in many practical applications. In many engineering practices, the system relative degree is larger than 1. The widespread presence of high relative degree dynamic systems has incited considerable interest in their ILC research within the control community in recent years.

For the high relative degree nonlinear continuous systems, under strict conditions of zero initial error, [12] adopted an antagonistic ILC method to enable system convergence; [13] designed a class of ILC algorithms based on data sampling; [14] proposed ILC laws which using error derivatives with the order less than the system relative degree. [15] presented a first-order D-type ILC based on the dummy model, which does not require the relative degree to be known. When there is a fixed initial shift, the control law proposed in [16] achieved consistent tracking over a specified interval by incorporating an initial correction behavior. This correction strategy had also been used in [17], [18] for high relative degree nonlinear discrete-time systems with fixed initial states.

For the high relative degree linear continuous systems, when the initial error is 0, [19] proposed a linear matrix inequality(LMI) design method based on the bounded real lemma(BRL). The research in [20] presented a unified 2-D analysis method for both continuous and discrete-time systems by defining similar symbols for continuous and discrete operators, but the model can only achieve asymptotic tracking for systems with initial shifts. The study in [21] proposed a PD-type control law for the fixed initial shifts, which guarantees convergence of the system within a finite interval, and the convergence speed is uniform. For uncertain systems with fixed initial shifts, [22] proposed an adaptive ILC algorithm. For the high relative degree linear multi-variable discrete-time systems, [23] proposed an iterative learning controller with an H_∞-based approach to suppress the random iteration-varying perturbations; when the system has a fixed iteration initial error, a P-type ILC algorithm is presented in [24] can achieve asymptotic tracking.

Regarding the initial value problem of ILC, most of the studies require that the initial shift is zero or fixed [25]-[27]. However, in many practical situations, the system will always inevitably exist initial shifts at each iteration, and due to the limitation of the actual repeat localization accuracy, the study of ILC with arbitrary initial shifts is of great significance. For nonlinear systems with varying initial iteration errors and tracking trajectories, the study in [28] proposed two adaptive ILC laws to achieve a complete reference trajectory tracking. The study in [29] presented a ILC method with a time-varying sliding mode, which enables random initial state errors to converge to zero beyond a initial time interval. This strategy achieves complete tracking for second-order nonlinear systems. An adaptive ILC algorithm based on filtering error correction is proposed in [30] to achieve precise tracking for non-parametric uncertain systems with random initial shifts and unknown input dead zones. For linear discrete time-delay systems, [31] proposed an ILC strategy with correction of initial state deviation to solve the trajectory tracking problem. The research in [32] adopted a phased ILC strategy, which first corrected arbitrary initial deviation to fixed deviation, and then corrected the fixed deviation, to achieve complete tracking for second-order continuous systems.

Despite these remarkable advances on ILC for arbitrary initial value problems, the study of high relative degree systems with arbitrary initial shifts is still relatively scarce. For the linear continuous MIMO systems with vector relative degree, the study in [33] proposed a control strategy based on an iteratively moving average operator, which make the system converge under a fixed initial error condition. Under an arbitrary initial error condition, this algorithm only made the system converge to a bounded range. For high relative degree SISO continuous systems, the presented ILC algorithm in [34] is based on the high relative degree, which utilized a form of multi-pulse compensation to suppress arbitrary initial shifts. For high relative degree linear discrete-time MIMO systems, [35] presented three ILC algorithms based on average operator to achieve complete tracking under the condition that the initial state vibrate slightly near a fixed point.

In this paper, an ILC algorithm based on compression mapping are presented to solve the random initial shift problem for high relative degree linear discrete-time systems. This algorithm draws on the phased correction strategy in [32] to deal with the random initial deviation problem. The random initial shifts are transformed into a fixed shift using a difference controller, and then the fixed shift is corrected by a PD-type controller to make the system converge at a predefined time. Finally, the validity of the proposed algorithm is demonstrated by simulation of two examples with different relative degrees. The conclusions are presented.

II. Problem Formulation

Consider a linear discrete-time system operating in the interval $[0, T]$:

$$\begin{aligned} x_k(t+1) &= Ax_k(t) + Bu_k(t) \\ y_k(t) &= Cx_k(t) \end{aligned} \tag{1}$$

where, $k = 1, 2, \cdots$ denotes the number of iterations; $x_k(t) \in \mathbb{R}^n$, $u_k(t) \in \mathbb{R}^n$, $y_k(t) \in \mathbb{R}^n$ denote the state variable, input and output of the system, respectively. A,B,C are system parameter matrices, which B is right invertible and C is left invertible.

$y_d(t)$ is the given desired output, $x_d(t)$ is the corresponding desired state. The system output error is defined as follows:

$$e_k(t) = y_d(t) - y_k(t) \tag{2}$$

Definition 1: [35] For the linear discrete-time system (1), if the Markov parameters satisfy

$$\left\{ \begin{array}{ll} CA^iB = 0, & 0 \leq i \leq p-2 \\ CA^{p-1}B \neq 0. & \end{array} \right. \tag{3}$$

The system relative degree of system is p.

Assumption 1: The initial state $x_k(0) \neq x_d(0)$ varies only arbitrarily within a certain range, i.e., $x_k(0)$ is a neighborhood of $x_d(0)$.

$$D = \{x_k(0) | \mid x_k(0) - x_d(0) \mid \leq \tfrac{\Lambda}{2}, x_k(0) \neq x_d(0)\}$$

where, $\frac{\Lambda}{2}$ is the radius of the neighborhood D.

III. Controller Design

In order to correct the random initial shifts of system (1), the controller is designed as follows:

$$\begin{aligned} & u_{k+1}(t) \\ = \; & u_k(t) + \textstyle\sum_{i=0}^{p-1} K_i(e_{k+1}(t-i) - e_k(t-i)) + r_k(t) \end{aligned} \tag{4}$$

where, K_i are control gains that can be set manually. $r_k(t)$ is a undetermined function.

Considering $x_{k+1}(t+p) - x_k(t+p)$, there is

$$\begin{aligned} & x_{k+1}(t+p) - x_k(t+p) \\ = \; & A(x_{k+1}(t+p-1) - x_k(t+p-1)) \\ & +B(u_{k+1}(t+p-1) - u_k(t+p-1)) \end{aligned} \tag{5}$$

Combining the control law (4) and Eq. (5), there is

$$\begin{aligned} & x_{k+1}(t+p) - x_k(t+p) \\ = \; & A(x_{k+1}(t+p-1) - x_k(t+p-1)) \\ & +B \textstyle\sum_{i=0}^{p-1} K_i(e_{k+1}(t+p-1-i) - e_k(t+p-1-i)) \\ & +Br_k(t+p-1) \\ = \; & A(x_{k+1}(t+p-1) - x_k(t+p-1)) \\ & -B \textstyle\sum_{i=0}^{p-1} K_iC(x_{k+1}(t+p-1-i) - x_k(t+p-1-i)) \\ & +Br_k(t+p-1) \end{aligned}$$

Setting $\eta_k(t) = x_{k+1}(t) - x_k(t)$, there is

$$\begin{aligned} & \eta_k(t+p) + (BK_0C - A)\eta_k(t+p-1) \\ & +B \textstyle\sum_{i=1}^{p-1} K_iC\eta_k(t+p-1-i) \\ = \; & Br_k(t+p-1) \end{aligned} \tag{6}$$

Eq. (6) is a p-order linear non homogeneous difference equation with constant coefficients. The corresponding homogeneous difference equation is

$$\begin{aligned} & \eta_k(t+p) + (BK_0C - A)\eta_k(t+p-1) \\ & +B \textstyle\sum_{i=1}^{p-1} K_iC\eta_k(t+p-1-i) = 0 \end{aligned} \tag{7}$$

Let its general solution be as follows:

$$\eta_k(t) = \varphi_k(t) + \varphi_k^*(t)$$

where $\varphi_k(t)$ is the general solution of Eq. (7) and $\varphi_k^*(t)$ is a particular solution of Eq. (6). One can set $\varphi_k(t) = \sum_{j=1}^{p} C_j\lambda_j^t$. λ_j are the p characteristic roots of the characteristic Eq. $\lambda^p + (BK_0C - A)\lambda^{p-1} + B\sum_{i=1}^{p-1} K_iC\lambda^{p-1-i} = 0$; C_j are arbitrary constant matrices.

Let characteristic equation have only one root λ, there is

$$\lambda = \lambda_j = \tfrac{A - BK_0C}{p} \tag{8}$$

The control gains K_i $(i = 1, 2, \cdots, p-1)$ can be obtained as follows:

$$K_i = B^{-1}C_p^{i+1}(-\lambda)^{i+1}C^{-1} = \tfrac{B^{-1}p!(-\lambda)^{i+1}C^{-1}}{(p-i-1)!(i+1)!} \tag{9}$$

Then $\varphi_k(t)$ can be written as:

$$\varphi_k(t) = \textstyle\sum_{s=0}^{p-1} C_st^s\lambda^t$$

For convenience, one can set $r_k(t+p-1) = Q\lambda^t$, where Q is an undetermined constant matrix. There is $r_k(t) = Q\lambda^{t-p+1}$. There is

$$\varphi_k^*(t) = qt^p\lambda^t$$

where q is an undetermined constant matrix.

Substituting the particular solution $\varphi_k^*(t)$ and Eq. (9) into Eq. (5), there is

$$\begin{array}{rl} & BQ\lambda^t \\ = & q(t+p)^p\lambda^{t+p} + (BK_0C - A)q(t+p-1)^p\lambda^{t+p-1} \\ & +B\sum_{i=1}^{p-1} K_iC(t+p-i-1)^p q\lambda^{t+p-i-1} \\ = & ((t+p)^p - p(t+p-1)^p \\ & +\sum_{i=1}^{p-1} C_p^{i+1}(-1)^{i+1}(t+p-i-1)^p)q\lambda^{t+p} \\ = & \sum_{j=0}^{p} C_p^j(-1)^j(t+p-j)^p q\lambda^{t+p} \\ = & \sum_{j=0}^{p} C_p^j(-1)^j(-j)^p q\lambda^{t+p} \end{array} \tag{10}$$

Thus

$$Q = B^{-1}\sum_{j=0}^{p} C_p^j(-1)^j(-j)^p q\lambda^p \tag{11}$$

$\eta_k(t)$ can be represented as

$$\begin{array}{rl} & \eta_k(t) = \sum_{s=0}^{p-1} C_s t^s\lambda^t + qt^p\lambda^t \\ = & \sum_{s=0}^{p} C_s t^s\lambda^t \end{array} \tag{12}$$

where $C_p = q$.

When $t \in [0, p-1]$, there are

$$\begin{cases} \eta_k(0) = C_0 \\ \eta_k(1) = \sum_{s=0}^{p} C_s\lambda \\ \eta_k(2) = \sum_{s=0}^{p} C_s 2^s\lambda^2 \\ \vdots \\ \eta_k(p-1) = \sum_{s=0}^{p} C_s(p-1)^s\lambda^{p-1} \end{cases} \tag{13}$$

If at a certain time $t = h$, there is $\eta_k(h) \to 0$, thus

$$\eta_k(h) = \sum_{s=0}^{p} C_s h^s\lambda^h = 0 \tag{14}$$

From (13) and (14), the coefficients C_s in (12) can be obtained as follows.

$$\begin{pmatrix} C_0 \\ C_1 \\ C_2 \\ \vdots \\ C_{p-1} \\ C_p \end{pmatrix} = \Upsilon \begin{pmatrix} \eta_k(0) \\ \eta_k(1) \\ \eta_k(2) \\ \vdots \\ \eta_k(p-1) \\ 0 \end{pmatrix} \tag{15}$$

where,

$$\Upsilon = \begin{pmatrix} I & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \lambda & \lambda & \lambda & \cdots & \lambda & \lambda \\ \lambda^2 & 2\lambda^2 & 2^2\lambda^2 & \cdots & 2^{p-1}\lambda^2 & 2^p\lambda^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda^{p-2} & (p-2)\lambda^{p-2} & (p-2)^2\lambda^{p-2} & \cdots & (p-2)^{p-1}\lambda^{p-2} & (p-2)^p\lambda^{p-2} \\ \lambda^{p-1} & (p-1)\lambda^{p-1} & (p-1)^2\lambda^{p-1} & \cdots & (p-1)^{p-1}\lambda^{p-1} & (p-1)^p\lambda^{p-1} \\ \lambda^h & h\lambda^h & h^2\lambda^h & \cdots & h^{p-1}\lambda^h & h^p\lambda^h \end{pmatrix}^{-1}$$

When $t = h \ (h > p)$ and $\eta_k(h) \to 0$, there are $x_{k+1}(h) \to x_k(h)$, $x_k(h) \to x_{k-1}(h)$, $\cdots$, $x_3(h) \to x_2(h)$,. That is, when $t = h$, the system output $y_{k+1}(h)$ tends to a certain steady value, but not necessarily $y_d(h)$.

In order to converge the output error to zero, the controller is modified as follows:

$$u_{k+1}(t) = \begin{cases} u_k(t) + \sum_{i=0}^{p-1} K_i(e_{k+1}(t-i) - e_k(t-i)) & \\ +r_k(t) & t \in [0,h] \\ u_k(t) + K_d(e_{k+1}(t) - e_k(t)) & \\ +\Gamma e_k(t+p) & t \in (h,T] \end{cases} \tag{16}$$

where K_d and Γ are control gains that can be set manually.

IV. Convergence Analysis

This section focuses on he convergence analysis of the system (1) after applying the control law (16). In this section, $||\cdot||$ is defined to be a certain norm for vectors or matrices.

Theorem 1: When the initial shifts satisfies **Assumption 1**, and the control gain Γ satisfies

$$||I - B\Gamma CA^{p-1}|| < 1 \tag{17}$$

Then the correction control law (16) can make the system (1) achieve complete tracking,that is $\lim_{k\to\infty}||e_{k+1}(t)|| = 0$.

Proof: According to the **Definition 1**, when $t \in [0,h]$, $CA^{i-1}B = 0$, one has

$$\begin{array}{rl} & y_k(t+i) \\ = & Cx_k(t+i) \\ = & CAx_k(t+i-1) + CBu_k(t+i-1) \\ = & CA^2x_k(t+i-2) + CABu_k(t+i-2) \\ & +CBu_k(t+i-1) \\ = & CA^ix_k(t) + \sum_{j=1}^{i} CA^{j-1}Bu_k(t+i-j) \\ = & CA^ix_k(t) \end{array} \tag{18}$$

In the analysis of the previous section, through the correction of the control law (16), the output error of the system (1) is stabilized at a fixed value when $t = h$.

When $t > h$, one has

$$\begin{array}{rl} & x_{k+1}(t) - x_k(t) \\ = & A(x_{k+1}(t-1) - x_k(t-1)) \\ & +BK_d(e_{k+1}(t-1) - e_k(t-1)) \\ & +B\Gamma e_k(t+p-1) \end{array} \tag{19}$$

From the Eq. (18), there is

$$e_k(t+p-1) = CA^{p-1}(x_d(t) - x_k(t)) \tag{20}$$

Setting $\Delta x_k(t) = x_d(t) - x_k(t)$. Substituting $\Delta x_k(t)$ into (19), one has

$$\begin{array}{rl} & \Delta x_k(t) - \Delta x_{k+1}(t) \\ = & (A - BK_dC)(\Delta x_k(t-1) - \Delta x_{k+1}(t-1)) \\ & +B\Gamma CA^{p-1}\Delta x_k(t) \end{array} \tag{21}$$

When $t = h+1$, according to $x_k(h) = x_{k+1}(h)$ and (21), there is

$$\Delta x_{k+1}(h+1) = (I - B\Gamma CA^{p-1})\Delta x_k(h+1)$$

Therefore, when $||I - B\Gamma CA^{p-1}|| < 1$, there is

$$\lim_{k\to\infty} ||\Delta x_{k+1}(h+1)|| = 0 \tag{22}$$

So

$$\begin{aligned} & \lim_{k\to\infty} ||e_{k+1}(h+1)|| \\ = \; & C\lim_{k\to\infty} ||\Delta x_{k+1}(h+1)|| \\ = \; & 0 \end{aligned}$$

Similarly, when $t \in (h+1, T]$, there is

$$\lim_{k\to\infty} ||e_{k+1}(t)|| = 0$$

V. Numerical Simulations

To verify the above conclusion, this paper conducts simulation experiments on two systems with relative degree 1 and 2, respectively. And compared it with the algorithm proposed in [24].

A. Linear Discrete-time MIMO System with Relative Degree 1

Considering the following MIMO system:

$$\begin{aligned} x_k(t+1) &= \begin{bmatrix} 0.6 & 0.25 \\ 0 & 0.65 \end{bmatrix} x_k(t) + \begin{bmatrix} 1 & 0.05 \\ 0.1 & 1.7 \end{bmatrix} u_k(t) \\ y_k(t) &= \begin{bmatrix} 0.8 & -0.13 \\ 0.1 & 0.5 \end{bmatrix} x_k(t) \end{aligned} \tag{23}$$

From $CB \neq 0$, the system relative degree is $p = 1$. Let the control gains of (16) be

$$K_0 = 1.2, \qquad K_d = 0.9, \qquad \Gamma = \begin{bmatrix} 0.27 & 0.14 \\ 0.03 & 0.31 \end{bmatrix}$$

It is easy to verify that $||I - B\Gamma CA^{p-1}|| < 1$, which satisfies the condition of **Definition 1**. $r_k(t)$ is determined according to Eq. (11) and (15). The system reference trajectory is as follows:

$$y_d(t) = \begin{bmatrix} y_{1,d}(t) \\ y_{2,d}(t) \end{bmatrix} = \begin{bmatrix} 0.0008(t-50)^2 - 1 \\ sin(0.02\pi t) \end{bmatrix} \tag{24}$$

The system initial state is $x_k(0) = [rand + 0.5 \;\; rand - 0.5]^T$(where $rand$ generates a random value between 0 and 1). Let the system operating interval be $[0, 100]$, and the preset time $h = 15$. The simulation results are shown in Fig. $1-3$, where simulation is implemented for 50 iterations.

Fig. 1 shows the results of system outputs $y_{1,k}(t)$ and $y_{2,k}(t)$ tracking the reference trajectories $y_{1,d}(t)$ and $y_{2,d}(t)$ after different number of iterations, respectively. Fig. 2 shows the results of tracking errors $e_{1,k}(t)$ and $e_{2,k}(t)$. Fig. 3 shows the variations of system inputs $u_{1,k}(t)$ and $u_{2,k}(t)$. From Fig. $1-3$, it is obvious that when $t = 15$, the system outputs $y_{1,k}(t)$ and $y_{2,k}(t)$ tend to be fixed values, and when $t = 16$, both the system tracking errors are zero, the system achieves accurate tracking.

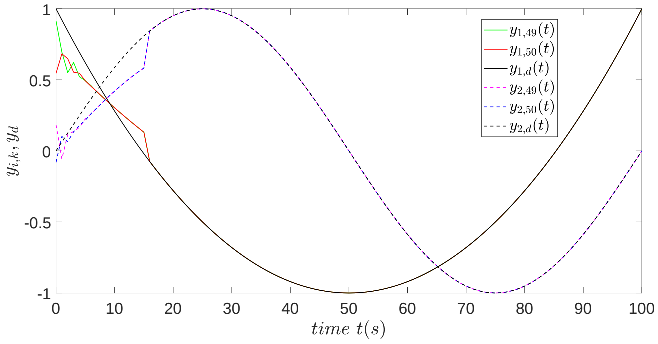


Fig. 1. Outputs $y_{i,k}(t)$ and reference trajectories $y_{i,d}(t)$.

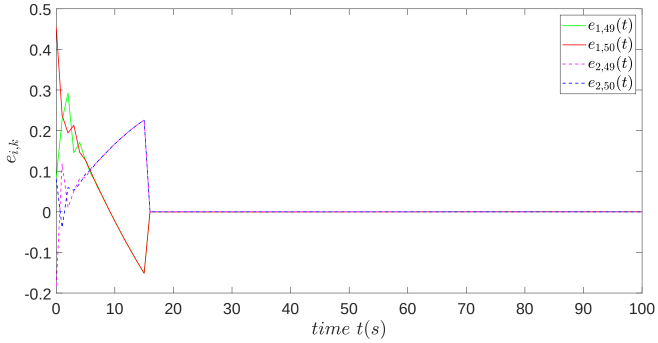


Fig. 2. Errors $e_{1,k}(t)e_{2,k}(t)$.

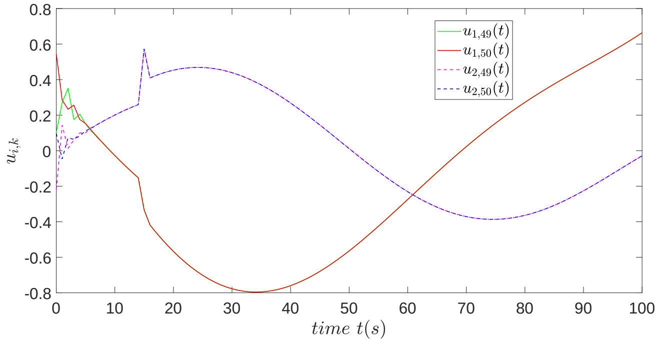


Fig. 3. Inputs $u_{1,k}(t)u_{2,k}(t)$.

B. Linear Discrete-time MISO System with Relative Degree 2

Considering the following MISO system:

$$\begin{cases} x_k(t+1) &= \begin{bmatrix} 0.8 & 0.1 \\ -0.25 & -0.33 \end{bmatrix} x_k(t) + \begin{bmatrix} 0 \\ 1.1 \end{bmatrix} u_k(t) \\ y_k(t) &= \begin{bmatrix} 1.7 & 0 \end{bmatrix} x_k(t) \end{cases} \tag{25}$$

From $CB = 0$ and $CAB \neq 0$, the system relative degree is $p = 2$. Let the control gains in the control law (16) be $K_0 = 1.8, K_d = 1.1, \Gamma = 2.9$. According to the relative degree $p = 2$ and Eq. (9), there exists $K_1 = -0.0052$. $r_k(t)$ is determined according to Eq. (11) and (15). It is also easy to verify that $||I - B\Gamma CA^{p-1}|| < 1$. Let the system operating interval be $[0, 100]$, and the preset time $h = 20$. The system desired trajectory is as follows:

$$y_d(t) = cos(0.02\pi t) \qquad t \in [1, 100]$$

The system initial state is $x_k(0) = [rand \quad rand]^T$. The simulation results are shown in Fig. $4-6$, where simulation is implemented for 50 iterations. Fig. 4 shows the result of the system output $y_k(t)$ tracking the desired trajectory $y_d(t)$ with the number of iterations. Fig. $5-6$ shows the variations of tracking error $e_k(t)$ and system input $u_k(t)$ after different number of iterations, respectively.

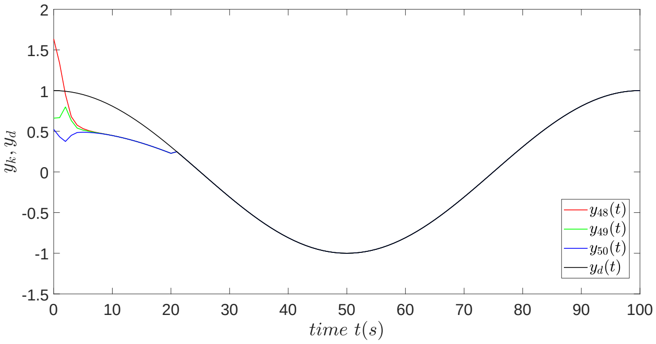


Fig. 4. $y_k(t)$ and $y_d(t)$.

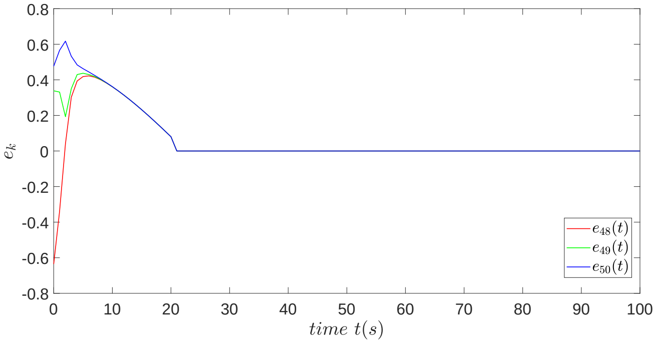


Fig. 5. Errors $e_k(t)$.

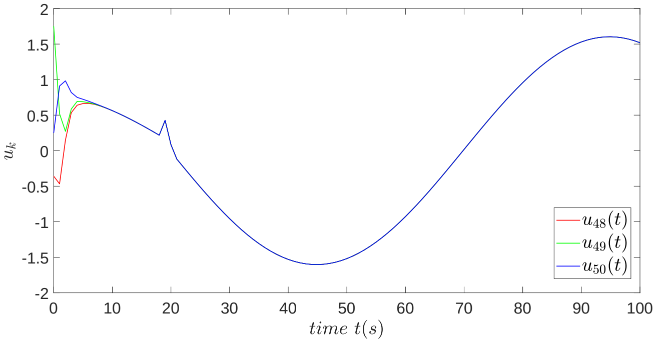


Fig. 6. Inputs $u_k(t)$.

From Fig. $4-6$, it is obvious that when $t = 20$, the system output tends to a fixed value. When $t = 21$, the system tracking error is 0, and the system achieves accurate tracking.

C. Comparison of Different Algorithms

For high relative degree linear discrete-time systems, the current research mainly focuses on the systems with fixed initial shifts, which is not applicable to the ones with random initial shifts. The ILC law (16) is compared with the algorithm proposed in [24] to demonstrate the effectiveness of the algorithm presented in this paper. Under the same conditions, the system (25) is simulated by using these two algorithms separately, and the results are shown in Fig. 7. The blue lines denote the outputs of the algorithm proposed in [24] after 48, 49, and 50 iterations respectively, and the black solid line denotes the desired trajectory, the red lines denote the outputs of the algorithm proposed in this paper. It is obvious that our algorithm can make the system converge, while the algorithm in [24] is unable to do.

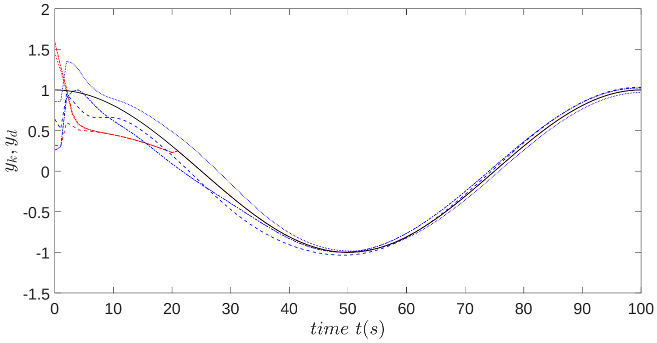


Fig. 7. $y_k(t)$ and $y_d(t)$.

VI. Conclusion

The ILC problem for high relative degree linear discrete-time systems with random initial shifts is discussed in this paper, and a control strategy with deviation correction is proposed. Theoretical analysis indicates that the presented algorithm can quickly correct the system initial state error and make the system converge after the predefined time. Finally, simulations were conducted using two discrete-time systems with relative degree 1 and 2, respectively, and the results proved the effectiveness of the algorithm. In the future, the effectiveness of this ILC strategy on discrete time-delay systems will be discussed.

Acknowledgment

This work was supported by the Beijing Orient Institute of Measurement and Test. [no. BOIMTSSHJD20182001].

References

[1] C. E. Boudjedir, M. Bouri and D. Boukhetala, "Model-Free Iterative Learning Control With Nonrepetitive Trajectories for Second-Order MIMO Nonlinear Systems Application to a Delta Robot," IEEE Transactions on Industrial Electronics, vol. 68, no. 8, pp. 7433-7443, Aug. 2021.

[2] S. Chen, S. Hsieh and T. Ta, "Iterative learning contouring control for five-axis machine tools and industrial robots," Mechatronics, vol. 94, pp. 103-030, Oct. 2023.

[3] D. X. Ba, N. T. Thien and J. Bae, "A Novel Iterative Second-Order Neural-Network Learning Control Approach for Robotic Manipulators," IEEE Access, vol. 11, pp. 58318-58332, 2023.

[4] V. Molazadeh, Q. Zhang, X. Bao and N. Sharma, "An Iterative Learning Controller for a Switched Cooperative Allocation Strategy During Sit-to-Stand Tasks with a Hybrid Exoskeleton," IEEE Transactions on Control Systems Technology, vol. 30, no. 3, pp. 1021-1036, May 2022.

[5] L. Liu, M. Illian, S. Leonhardt and B. J. E. Misgeld, "Iterative Learning Control for Cascaded Impedance-Controlled Compliant Exoskeleton With Adaptive Reaction to Spasticity," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-11, 2023.

[6] R. Hou, L. Cui, X. Bu and J. Yang, "Distributed formation control for multiple non-holonomic wheeled mobile robots with velocity constraint by using improved data-driven iterative learning," Applied Mathematics and Computation, vol. 395, pp. 125-829, Apr. 2021.

[7] X. Fu and J. Peng, "Iterative learning control for UAVs formation based on point-to-point trajectory update tracking," Mathematics and Computers in Simulation, vol. 209, pp. 1-15, July 2023.

[8] H. Shokri-Ghaleh, S. Ganjefar and A. M. Shahri, "Extension of iterative learning control design for batch processes with time-delay in the input subject to random cycle-varying uncertainties," Journal of the Franklin Institute, vol. 360, no. 12, pp. 8528-8549, Dec. 2023.

[9] H. Li, S. Wang, H. Shi, C. Su and P. Li, "Two-Dimensional Iterative Learning Robust Asynchronous Switching Predictive Control for Multiphase Batch Processes With Time-Varying Delays," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, no. 10, pp. 6488-6502, Oct. 2023.

[10] Y. Chen, D. Huang, Y. Li and X. Feng, "A Novel Iterative Learning Approach for Tracking Control of High-Speed Trains Subject to Unknown Time-Varying Delay," IEEE Transactions on Automation Science and Engineering, vol. 19, no. 1, pp. 113-121, Jan. 2022.

[11] J. Zheng and Z. Hou, "Data-Driven Spatial Adaptive Terminal Iterative Learning Predictive Control for Automatic Stop Control of Subway Train With Actuator Saturation," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 10, pp. 11453-11465, Oct. 2022.

[12] M. Sun and D. Wang, "Anticipatory Iterative Learning Control for Nonlinear Systems with Arbitrary Relative Degree," IEEE Transactions on automatic control, vol. 46, no. 5, pp. 783-788, 2001.

[13] M. Sun and D. Wang, "Sample-data iterative learning control for nonlinear systems with arbitrary relative degree," Automatic, vol. 37, no. 2, pp. 283-289, 2001.

[14] M. Sun and D. Wang, "Higher relative degree nonlinear systems with ILC using lower-order differentiations," Asian Journal of Control, vol. 4, no. 1, pp. 38-48, 2002.

[15] Z. Song, J. Mao and S. Dai, "First-order D-type Iterative Learning Control for Nonlinear Systems with Unknown Relative Degree," Acta Automatica Sinica, vol. 31, no. 4, pp. 555-561, 2005.

[16] M. Sun and D. Wang, "Iterative learning control with initial rectifying action," Automatic, vol. 38, no. 7, pp. 1177-1182, 2002.

[17] M. Sun and D. Wang, "Analysis of nonlinear discrete-time systems with higher-order iterative learning control," Dynamics and Control, vol. 11, no. 1, pp. 81-96, 2001.

[18] M. Sun and D. Wang, "Initial shift issues on discrete-time iterative learning control with system relative degree," IEEE Trans. Autom. Control, vol. 48, no. 1, pp. 144-148, 2003.

[19] D. Meng, Y. Jia, Du J and F. Yu, "Monotonically convergent ILC systems designed using bounded real lemma," International Journal of Systems Science, vol. 43, no. 11, pp. 2062-2071, 2012.

[20] D. Meng, Y. Jia, Du J and F. Yu, "Data-driven control for relative degree systems via iterative learning," IEEE Transactions on Neural Networks, vol. 22, no. 12, pp. 2213-2225, 2011.

[21] Q. Fu, L. Du, G. Xu, J. Wu and P. Yu, "PD-type iterative learning control for linear continuous systems with arbitrary relative degree," Transactions of the Institute of Measurement and Control, vol. 41, no. 9, pp. 2555-2562, 2019.

[22] C. Chien and C. Yao, "An output-based adaptive iterative learning controller for high relative degree uncertain linear systems," Automatic, vol. 40, no. 1, pp. 145-153, 2004.

[23] D. Meng, Y. Jia, Du J and F. Yu, "Robust learning controller design for MIMO stochastic discrete-time systems: An H_∞-based approach," International Journal of Adaptive Control and Signal Processing, vol. 25, pp. 653-670, 2011.

[24] Z. Mao and X. Li, "Iterative learning control for linear discrete systems with high relative degree and iterative initial error," Control Theory & Applications, vol. 29, no. 8, pp. 1078-1081, 2012.

[25] D. Huang, J. Xu, X. Li X, et al, "D-type anticipator iterative learning control for a class in homogeneous heat equations," Automatica, vol. 49, pp. 2397-2408, 2013.

[26] X. Dai, C. Xu, S. Tian and Z. Li, "Iterative learning control for MIMO second-order hyperbolic distributed parameter systems with uncertainties," Advances in Difference Equations, vol. 94, pp. 1-13, 2016.

[27] P. GU, Q. FU and J. WU, "State Tracking Algorithm for Linear Singular Iterative Learning Control Systems with Fixed Initial Shift," Mathematica Applicata, vol. 30, no. 1, pp. 8-15, 2017.

[28] X. Li, M. Lv and K. John, "Adaptive ILC algorithms of nonlinear continuous systems with non-parametric uncertainties for non-repetitive trajectory tracking," International Journal of Systems Science, vol. 47, no. 10, pp. 2279-2289, 2016.

[29] C. Yin, S. Riaz, H. Zaman, N. Ullah, V. Blazek, L. Prokop and S. Misak, "A Novel Predefined Time PD-Type ILC Paradigm for Nonlinear Systems," mathematics, vol. 11, no. 1, 2023.

[30] Q. Yan, M. Sun and J. Cai, "Filtering-error rectified iterative learning control for systems with input dead-zone," Control Theory Applications, vol. 34, no. 1, pp. 77-84, 2017.

[31] G. Li, D. Chen, Q. Dong and K. Wang, "Iterative learning control for a class of discrete-time systems with time delay and random initial state errors," Pure and Applied Mathematics, vol. 38, no. 2, pp. 224-235, 2022.

[32] D. Chen, Y. Xu, T. Lu and G. Li, "Multi-phase iterative learning control for high-order systems with arbitrary initial shifts," Mathematics and Computers in Simulation, vol. 216, pp. 231-245, 2024.

[33] Y. Wei and X. Li, "Robust iterative learning control for linear continuous systems with vector relative degree under varying input trail lengths and random initial state shifts," International Journal of Robust and Nonlinear Control, vol. 31, no. 2, pp. 609-622, 2021.

[34] X. Ruan and J. Wang, "Iterative Learning Control for Linear Time-invariant Systems with Higher-order Relative Degree," Acta Mathematicae Applicatae Sinica, vol. 37, no. 6, pp. 1077-1092, 2014.

[35] Y. Wei and X. Li, "Iterative learning control for linear discrete-time systems with high relative degree under initial state vibration," IET Control Theory and Applications, vol. 10, no. 10, pp. 1115-1126, 2016.

A Memetic Algorithm to Solve the Two-Echelon Collaborative Multi-Centre Multi-Periodic Vehicle Routing Problem with Specific Constraints

Camelia Snoussi, Abdellah El Fallahi, Sarir Hicham
MaCS-DM Laboratory, National School of Applied Sciences
University AbdelMaleek Essaadi
Tetuan, Morocco

***Abstract*—The collaboration between distribution companies is gaining a great interest in the last years due to the benefit provided to reduce the cost of deliveries. In this work we study the centralized two-echelon collaborative multi-center multi-periodic vehicle routing problem with a specific constraints. In which each distribution center conserves its VIP customers, and each partner keep their delivery scheduling unchangeable. The problem is modelled as a MILP, and to solve it a hybrid algorithm is proposed. This algorithm combines a multi-population memetic algorithm (MPMA) and a variable neighbourhood search algorithm that integrates a tabu search list (VNS-T). The results obtained are compared with those obtained by CPLEX solver and the best known solution of the multi-depot vehicle routing problem (MDVRP).**

***Keywords*—Collaborative vehicle routing problem; two-echelon networks; memetic algorithm; Variable neighbourhood serach**

I. Introduction

The significant growth in the delivery of small volumes of goods generated by the increase in e-commerce sales [1] and domestic freight, particularly during the Covid-19 pandemic [2], creates major challenges for distribution companies operating in the urban sector. It's well known that the urban transport faces a major problem of empty runs with more than 40% of unloaded trips. The empty running of trucks generates multiple challenges like higher delivery costs [3], congestion of distribution networks [4], and increased CO2 emissions [5]. To face these challenges, distribution companies are driven to explore new distribution strategies.

Recently, centralized collaborative strategies have been gaining considerable attention due to their positive impact on the reduction of distribution costs. These strategies generate coalitions involving multiple independent members, in which the organization of the collaborative process can be outsourced to a third-party (TP). This collaboration is often described by the collaborative multi-center vehicle routing problem (CMCVRP) [6]. The maximum performance of a collaborative process is obtained through total information sharing between its members [7]. Nonetheless, the formation of new coalitions faces several challenges [8], especially in establishing the necessary level of trust between partners. The most common problem resides in the lack of background in the field of collaborative practices as well as an incomplete legislative framework [9]. Moreover, Companies are not willing to risk losing their Very Important (VIP) customers in favour of other coalition members. On the other hand, to prevent a possible deterioration of their service quality, companies try to keep their delivery schedules unchangeable.

To the best of our knowledge, there are no studies that tackle the centralized CMCVRP where the information related to VIP customers is concealed and the delivery schedule of each member of the coalition is kept unchanged. To fill this gap, this paper studies a new extension of the two-echelon collaborative multi-center vehicle routing problem (2E-CMCVRP) by considering constraints of VIP customers and inflexible delivery schedules [10]. The problem is formulated as a MILP, and a multi-phase solving approach (MPSA) is proposed to solve it. The MPSA integrates a multi-population memetic algorithm (MPMA) and a modified VNS algorithm.
The remainder of this paper is organized as follows. In Section II, the literature is reviewed. Then in Section III, we describe the problem and we formulate its corresponding mathematical model. In Section IV, a detailed description of the proposed multi-phase solving approach is presented. The numerical results are presented in the Section V. Finally, the conclusions and future research suggestions are presented in the Section VI.

II. Literature Review

A. Collaborative Two-echelon Periodic MCVRP

The underlying problem of this paper (CMCVRP) is an extension of the MDVRP. Several approaches to solve the MDVRP have been examined in the literature. In 2012 Vidal et al. proposed a hybrid genetic algorithm with an adaptive diversity control metaheuristic to solve the periodic MDVRP [11]. In 2015, Rahimi et al. introduced a new modular heuristic algorithm (MHA) to manage the periodic MDVRP with capacity, duration, and maximum budget constraints [12]. Recently, an extensive review of different formulations and solving approaches for the MDVRP was published by Ramos, Shara et al. [13], [14].

To deal with the different challenges, distribution companies are constrained to treat their competitors differently by introducing new forms of interaction. Therefore, various forms of collaboration are emerging in this sector that show an important potential benefits [15], [16], [17], [18], [19]. Moreover, different variants of the routing problem in a collaborative setting have been explored. In [20] the authors developed an adaptive large neighborhood search iterative algorithm to solve a pickup and delivery problem with time window (PDPTW) that considers the outsourcing and exchange of requests between collaborators in a centralized collaborative configuration. In 2016, Soysal et al. studied the impact of horizontal collaboration on perishable products, logistics costs, and CO2 emissions for the inventory routing problem (IRP) [21].

Generally, the urban distribution is a periodic VRP (PVRP) for which several studies focused on implementing collaborative scenarios are presented. In [22] an empirical study is conducted to evaluate the influence of three companies' characteristics (e.g., the number of orders to transport, the order size and the ability of a company to delay its orders) which operate in a periodic scenario on the total profit. In 2017, Smilowitz et al. developed an adaptive large neighborhood search algorithm to solve the periodic location routing problem in the collaborative recycling sector [23]. They concluded that increasing the flexibility of delivery schedules impacts positively the reduction of expenses, especially when the maximum capacity constraint is very stringent.

The two-echelon vehicle routing problem (2E-VRP) is a variant of the multi-echelon vehicle routing problem (MEVRP) where the delivery of goods to customers is done by various intermediate plants. The urban distribution is one of the sectors where the 2E-VRP is often implemented [24]. In [25] Wang et al. introduced the two-echelon collaborative multi-center vehicle routing problem (2E-CMCVRP) as a combination of a multi-center VRP (MCVRP) and a profit allocation problem. The authors evaluated the economic and environmental impact of such collaboration and they proposed a two-phase algorithm based on a clustering method and a non-dominated sorting genetic algorithm to solve it.

In 2020, Wang et al. [26] proposed a collaborative and resource-sharing strategy to solve the multi-depot multi-period vehicle routing problem with pickup and delivery (MD-PVRPPD). The authors concluded that combining collaborative and resource-sharing mechanisms improve the multi-depot multi-period logistics network with pickup and delivery.

B. Information Sharing in Decentralized and Centralized Collaboration Configurations

The development of the collaborative process involving different partners requires certain level of information sharing. Generally, there are two information sharing strategies: partial sharing in decentralized collaborative configurations and total sharing in centralized configurations [27]. In both strategies, a third-party (A logistics service provider, online platform...etc) it's responsible for organizing the collaboration.

The decentralized collaborative VRP has been the focus of several studies that considered different mechanisms of information sharing. In [28] the authors suggested a request exchange mechanism based on the auction of a single request with limited information. In 2017, Huang et al. [29] developed an efficient auction-based mechanism for the carrier collaboration problem with bilateral exchange (CCPBE) where the carriers can only offer requests with the highest marginal costs and can bid on a bundle of lanes. Under this mechanism, the shared lanes' information is available to all carriers. In [30], the authors evaluated the impact of providing information about requests to the auction pool in an auction-based carrier collaboration problem where the requests are shared in an aggregated form. The aggregates are generated through grids that cover the geographical area of the requests.

Different types of central authorities have been considered in studies investigating the centralized collaborative VRP. The dynamic collaborative pickup and delivery problem which relies on a peer-to-peer platform that matches ad hoc drivers or backup vehicles to deliver tasks in real time is introduced in [1]. In [17], Maneengam et al. developed the centralized collaborative bidirectional multi-period vehicle routing problem under profit-sharing agreements, where the collection and integration of information and resources are done by a control tower. The tower establishes a collaborative transport planning respecting the profit-sharing agreements. The integration of tactical collaborative decisions has been raised in [32], in which the authors evaluated the economic and environmental impact of two collaborative scenarios in a centralized configuration: semi-cooperative and fully cooperative. In the first scenario, collaboration takes place at the operational level where all the customers and vehicle capacity information are shared to build the routing plan. In the second scenario, the collaboration occurs not only on an operational level but also tactical one. In this scenario, the routing and facility location decisions are taken jointly.

III. The Two-echelon Collaborative Multi-center Multi-periodic Vehicle Routing Problem

A. Problem Description and Assumptions

The 2E-CMCPVRP with VIP customers and inflexible delivery schedules is a distribution network with several independent distribution centers (DC), in which we suppose that one of the centers has the needed infrastructure to play the role of a logistics center (LC). Each center i serves a set of customers according to a specific schedule w_i of several periods t with a fleet of K_i vehicles, some of these customers can be shared with other centers except its VIP customers. Furthermore, we consider that a neutral third-party is in charge of the organization of the centralized collaborative process whose objective is to establish a collaborative network by a possible reassignment of non-VIP customers. The volumes of goods corresponding to the reassigned customers are initially stored in the LC. The presence of a LC transforms the initially independent distribution network into a collaborative two-echelon system where the routes between the LC and the DCs are covered by semi-trailers. The major assumptions of this study are:

- The LC has enough additional storage space,
- Each center DC has enough storage space to accommodate the reassigned goods from other centers,
- The customer's demand is deterministic and is known a priori,
- Each customer is visited only once during the consolidated delivery schedule,
- The fleet of vehicles is homogenous and each DC has a limited number of vehicles,

- Each vehicle starts and ends its route at the center in which it is parked within a limited time,
- The semitrailer starts and ends its route at the LC,
- The average speed of the roads (arcs) may differ from one road to another,
- The maintenance, leasing, and fuel costs of the vehicles may differ from one vehicle to another.
- Model formulation

The proposed MILP model for the 2E-CMCPVRP-VCIS is formulated as a two-echelon collaborative periodic VRP model. The objective is to minimize the total distribution costs considering the VIP customers' constraints and the inflexible delivery schedules. The parameters and related notations used in the 2E-CMCPVRP-VCIS model are detailed in the next section.

B. Parameters and Notations

The parameters and notations used in this work are presented in Table I, Table II and Table III. The best values of the parameters are obtained by testing various values of each parameter.

TABLE I. Data Sets in the 2E-CMCPVRP-VCIS

Set	Definition
I	Set of centers (DC and LC)
J	Set of customers
K	Set of vehicles
T	Set of periods in w
K^i	Set of vehicles belonging to the center i, i∈ I
J^i	Set of customers initially assigned to center i, i∈ I

C. Modeling

a) Objective function: The objective function is defined as follows:

$$T_{total} = T_1 + T_2 + T_3 \tag{1}$$

where,T1 defines the sum of costs related to the semi-trailer's, T_2 is the sum of the dispatching, maintenance, leasing, and fixed costs and T_3 is the sum of delivery costs related to the fuel consumption during a consolidated delivery schedule.

$$T_1 = \sum_{t \in T} (T_{11}^t + T_{12}^t) \tag{2}$$

where, T_{11}^t and T_{12}^t are respectively the sum of the fuel costs and maintenance costs of the semi-trailer over a period *t* given by the Eq. (3) and Eq. (4).

$$T_{11}^t = \sum_{i,h \in I, h \neq i} H_{se} \times \rho \times d_{ih} \times o_{ih}^t \tag{3}$$

$$T_{12}^t = \sum_{(i,h \in I, h \neq i)} \frac{M \times o_{ih}^t \times d_{ih}}{K_a} \tag{4}$$

TABLE II. Input Parameters in the 2E-CMCPVRP-VCIS

Parameter	Definition
u_j^t	Equals 1 if the customer must be visited on period t, j∈J,t∈T
ρ	Fuel price
N_i	Number of vehicles belonging to the center i,i∈I
N_s	Number of semi-trailers
H_{se}	Average fuel consumption of the semi-trailer per 100km
h_k	Average fuel consumption of the vehicle k,k∈K per 100km
F	Average annual vehicle maintenance costs
L	Average annual vehicle rent or leasing costs
Q_{max}	Maximum capacity of a vehicle
N_{Tot}	Total number of available vehicles
K_a	Average annual distance covered by a semi-trailer (Km)
B	The capacity of the semi-trailer
T_{max}	Maximum working time per period
q_j	The demand of the customer j,j∈ J
M	Average annual semi-trailer maintenance costs
G_i	Fixed costs of center i per period. The third-party (TP) covers the fixed costs when the centre i agrees to cooperate, i∈I
P_i	CA's service costs for center i per period when cooperation is achieved, i∈I
τ	Number of consolidated delivery schedules per year
w_i	Delivery schedule of center i,i∈I
w	Consolidated delivery schedule of the coalition with $w = \cup_{i \in I} w_i$
d_{ij}	The distance between centers i and j, i,j∈ $J \cup I$
y_i	Coefficient of variable costs of center i, i∈I
vip_{ij}	If the customer j is a VIP customer of centre i, $vip_{ij} = 1$ else $vip_{ij} = 0$,j∈J, i∈I
v_{ij}	Average road speed between nodes i and j, (j,i)∈ $J \cup I$
V_{ik}	Assignment of a vehicle to a specific center, i∈I,k∈K
y_i	y_i=1 if centre i collaborates else y_i= 0, i∈I

TABLE III. Decision Variables

Variable	Definition
x_{ij}^{kt}	Equals 1 if vehicle k travels directly from i to j during the period t otherwise is equal to 0, (i,j)∈I∪J,k∈K,t∈T
o_{ij}^t	If the semi-trailer travels directly from center i to centre j on period t, $o_{ij}^t = 1$ else $o_{ij}^t = 0$, i,j∈I,t∈T
ϕ_{ik}	A variable used for the elimination of sub-turns in the second echelon. It is always positive, i∈I,k∈K
δ_{ik}	A variable used for the elimination of the sub-turns in the first echelon. It is always positive, i∈I,k∈K

T_2 gives the sum of the dispatching, maintenance, leasing, and fixed costs as in the Eq. (5).

$$T_2 = T_{21} + \sum_{t \in T} T_{22}^t. \tag{5}$$

where T_{21} Gives the total service costs required by the third-party plus the maintenance and leasing costs of the fleet, and T_{22}^t is the dispatching costs of the quantities delivered during a period *t*. where:

$$T_{21} = \sum_{i \in I} [(1 - y_i)G_i + y_i P_i + (T_{23} \times (\frac{F + L}{\tau}))] \tag{6}$$

where T_{23} gives the needed number of vehicles to cover the customers' demands over the consolidated delivery schedule; it is equal to the highest number of vehicles used by all centers during a period *t*.

$$T_{22}^t = \sum_{i \in I} \sum_{k \in K^i} \sum_{p \in I \cup J} \sum_{j \in J} x_{ij}^{kt} \times u_j^t \times q_j \times \gamma_i. \tag{7}$$

T_3 is the sum of delivery costs related to the fuel consumption during a consolidated delivery schedule as in Eq. (8)

$$T_3 = \sum_{t \in T} \sum_{i,j \in I \cup J} \sum_{k \in K} \frac{d_{ij} \times x_{ij}^{kt} \times u_j^t \times \rho \times h_k}{100}. \quad (8)$$

First echelon constraints:

$$\sum_{j \in J, j \neq i} o_{ji}^t = 1, i \in I, t \in T (i = 1 \text{ corresponds to LC}) \quad (9)$$

$$\sum_{i \in J, j \neq i} o_{ij}^t = 1, j \in I, t \in T (j = 1 \text{ corresponds to LC}) \quad (10)$$

$$\sum_{j \in I} o_{ij}^t - \sum_{j \in I} o_{ji}^t = 0, i \in I, i \neq, t \in T \quad (11)$$

$$\sum_{i \in I} \sum_{k \in K_i} \sum_{l \in I \cup J, p \in J \setminus J^i, l \neq i)} x_{ij}^{kt} \times u_j^t \times q_p \leq B, t \in T \quad (12)$$

$$\phi_i - \phi_j + N_s \times o_{ij}^t \leq (N_s - 1), i, j \in I, i \neq 1, t \in T \quad (13)$$

$$\phi_i \geq 0, i \in I, i \neq 1 \quad (14)$$

Second echlon constraints:

$$\sum_{t \in T} \sum_{k \in K} \sum_{i \in I \cup J, i \neq j} x_{ij}^{kt} \times u_j^t = 1, j \in J \ (15)$$

$$\sum_{j \in J} (q_j \times \sum_{i \in I \cup J} x_{ij}^{kt}) \leq Q_{max}, k \in K, t \in T \ (16)$$

$$\delta_{ik} - \delta_{jk} + N_v \times x_{ij}^{kt} \leq N_{Tot} - 1, k \in K, i, j \in J, t \in T \ (17)$$

$$\delta_{ik} \geq 0, k \in K, i \in J \ (18)$$

$$\sum_{j \in I \cup J} x_{ij}^{kt} - \sum_{j \in I \cup J} x_{ji}^{kt} = 0, k \in K, i \in I \cup J, t \in T \ (19)$$

$$\sum_{i,j \in I \cup J} x_{ij}^{kt} \times \frac{d_{ij}}{v_{ij}} \leq T_{max}, k \in K, t \in T \ (20)$$

$$\sum_{t \in T} \sum_{K^p} \sum_{i \in I \cup J} x_{ij}^{kt} \geq vip_i j, p \in I, j \in J \ (21)$$

$$\sum_{k \in K} \sum_{j \in J} x_{ij}^{kt} \leq N_i, i \in I, t \in T \ (22)$$

$$\sum_{J \in J} x_{ij}^{kt} - V_{ik} = 0, i \in I, k \in K, t \in T \ (23)$$

$$T_{23} \geq \sum_{k \in K} \sum_{i \in I, p \in J} x_{ij}^{kt} \ (24)$$

$$x_{ij}^{kt} \in \{0, 1\}, i \in I \cup J, j \in J, k \in K, t \in T \ (25)$$

$$\theta_{ij}^t \in \{0, 1\}, i \in I, j \in J, t \in T \ (26)$$

Constraints (15) ensure that each customer must be visited only once during the consolidated delivery schedule. Constraints (16) concern the vehicle's capacity. Constraints (17) and (18) are used for the vehicle's sub-tours elimination. Constraints (19) guarantee the flow conservation from/to each customer. Constraints (20) limit the vehicle travel time. Constraints (21) state that if j is a VIP customer it can be served only if there is a vehicle starting from its original center, otherwise, if it is not a VIP customer, it can be served by any available vehicle. Constraints (22) limit the number of vehicles starting from a center to the number of vehicles belonging to this center. Constraints (23) guarantee that any vehicle that starts from a center must belong to this center. Constraints (24) ensure that the number of vehicles to cover all the customer requests throughout the consolidated delivery schedule equals the sum of the maximum of vehicles used by each center in the busiest period. Constraints (25) and (26) ensure that the decision variables are binary.

IV. Multi-phase Solution Approach

The main objective of this approach is to determine the optimal coalition that minimizes the distribution cost and assures the best individual profit for its members. The approach is divided into two phases. In the first phase, we optimize the second echelon routes using a multi-population memetic algorithm (MPMA). And, in the second phase, to optimize the semi-trailer's routes we propose a variable neighbourhood search (VNS) algorithm that integrates a tabu list mechanism. These two phases are performed for each period of the delivery schedule.

Memetic algorithms are a hybridization of a genetic algorithm (GA) with local search heuristics. They are widely adopted in the resolution of routing problems [33], [34]. In the proposed MPMA, the chromosomes are encoded as a giant tour as presented in sub-section IV-A. The solutions are firstly evaluated using the clustering algorithm detailed in sub-section IV-B, and secondly by an improved splitting algorithm given by the pseudo-code 2. The MPMA uses three same-sized populations to avoid premature convergence [35]. Two populations are relaxed and may contain infeasible solutions P_{relax1} and P_{relax2} while the third one contains only feasible solutions$P_{feasible}$. To build the initial populations a clustering method is used as described in sub-section IV-C.

In P_{relax1} the fleet size N_i is relaxed according to the following equation $N_i = CR_v \times N_i, i \in I, CR_v \geq 1$ while in P_{relax2}, the maximum working time T_{max} and the maximum capacity Q_v are relaxed as follows $T_{max} = CR_t \times T_{max}$ and $Q_v = CR_l \times Q_v, CR_t \geq 1, CR_l \geq 1$ where CR_v is the vehicles number relaxation coefficient, CR_t is the maximum route duration relaxation coefficient and CR_l is the maximum load relaxation coefficient. During the search process, the populations remain sorted in ascending order according to their solutions' fitness values. In each generation, the algorithm selects two parent solutions from each population and then performs the crossover procedure presented in sub-section IV-D. The resulting offspring solutions are then evaluated using the modified Beasley-Bellman algorithm described in sub-section IV-G. If the new solution is feasible it is inserted in $P_{feasible}$ if it verifies the insertion conditions. Otherwise, the new solution will be inserted in one of the corresponding unfeasible populations if it checks the insertion conditions. The insertion conditions are: (1) the new solution should be different from all the existing solutions in the population and (2) the new solution should outperform the worst one in this population. To intensify the search around the newly generated solutions, a local search procedure, as presented in sub-section (e), is performed after each $n = Freq_{loc}$ generations. The algorithm then calculates the second echelon costs T_3 and based on the best feasible solution it calculates T_{22}. Then, the first echelon route optimization is performed as detailed in sub-section (g). When all the periods are processed the MPMA calculates T_{23} and T_{21}. Finally, the sub-coalition total cost is computed.

The pseudo-code of MPMA is shown in Algorithm 1.

Algorithm 1 Multi-phase solving approach

```
Load data
for <each sub-coalition Of industries> do
    Establish the consolidated schedule based on T
    for each period Of T do
        for each population do
            Choose the clustering parameters $N_{initial}, wf, wd, Max_{deviation}$
            Create a giant tour for each DC
            Determine the intersection zones between DCs
            Relax $T_{max}$ and split the giant tour into trips
            Perform the route sequencing
            for n fro 1 to $n_{max}$ do
                Perform the intra-routes Swap(1,1), inter-routes Swap(i,i)
                Insert the generated solutions into the current population
                $n = n + 1$
            end for
        end for
    end for
    for g→1 to $g_{max}$ do
        if counter=Freqloc then then
            Perform the VNS-Tabu search algorithm
            Perform the diversification heuristic
        end if
        Perform the parents' selection procedure
        Perform the crossover procedure
        Perform the advanced split procedure
        if Is Feasible Population then
            if Is a Feasible Solution then
                Compute cost
                Insert offspring into the feasible population according to the cost order
                if Is New Best Solution then
                    Update the new best solution
                    Insert into the relaxed population according to the penalized cost order
                end if
            else
                Compute penalized cost
                Insert into the relaxed population according to the penalized cost order
                Remove the worst solution from the population
                $g = g + 1$
            end if
        end if
    end for
    With the best solution from the feasible population
    for g → 1 to g do
        Compute the second echelon routes costs
        Compute goods' exchange between collaborating DCs in period w
        Compute used vehicles per DC
        Improve semitrailer routes
        Compute semitrailer routes costs
        Determine necessary vehicles per DC
        Compute the maintenance and leasing vehicles' costs
        Compute the centers fixed and variables costs
        Return the total costs for each sub-coalition
    end for
end for
```

A. Chromosome encoding

The chromosome encoding is defined by a giant tour divided into routes by delimiters representing the center from which each route starts. The node numbering $(0, \ldots, i-1)$ represents the centres and the following numbers $(i, \ldots, n+i-1)$ represent the set of n customers. During the crossover, the route delimiters are removed and then reinserted later using the splitting algorithm as shown in Fig. 1.

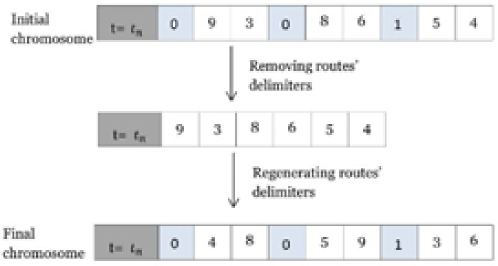


Fig. 1. Chromosome encoding in the MPMA.

B. Chromosome Evaluation

To build the solution corresponding to each chromosome, we apply the splitting procedure described in Algorithm 2. This algorithm is inspired by Vidal's adapted version of the Beasley-Bellman method [36] and integrates specific constraints of our model. During the splitting process, we perform an extraction of intermediate solutions while updating the set of the needed vehicles for each centre. The feasible sub-sequences of customers $T_{ts} = (T_t, \ldots, T_s), t \in [0, n], s \in [t+1, n]$ are evaluated using two nested loops as shown in Algorithm 2. The duration of the route $(dc, \sigma T_{ts}, dc)$, where σT_{ts} is a circular permutation of T_{ts}, is calculated by choosing the dc having available free vehicles and offering minimal service time. If among the customers belonging to sub-sequence Tts there is a VIP customer of a centre i then this centre must be the starting and ending node of the route.

C. Generation of Initial Population

To generate the initial population, we use the Split Middle Line Clustering (SMLC) method, which is a variant of the route-first cluster-second method. The SMLC method takes into consideration the collaborative and periodic aspects of the problem.

a) Step1: Initial clustering: This procedure creates a giant tour or a cluster cl_{it} for each pair center-period (i, t), which includes the centers' VIP customers and closest customers as shown in Fig. 2(b). After that, it creates shared zones between each pair of clusters cl_{it} and cl_{jt} which include non-VIP customers for whom the distance to the nearest customers belonging to the other cluster is smaller than the distance to the nearest customers of their own cluster (see Fig. 2(b)).

b) Step2: Route splitting: The splitting procedure divides the giant tour into routes, respecting the problem's constraints, by assigning customers from the center's giant tour and shared areas to the routes as follows:

- Place the center dc_i as the first element of each route r_i^m starting from dc_i where m is the index of the route and i is the number of the centre.
- Add the closest node n_i to dc_i as the second element of r_i^m and determine the barycentre z of (dc_i, n_i).
- Add the closest node n_2 to the barycentre and determine the new barycentre of (dc_i, n_1, n_2). Repeat this

Algorithm 2 : Splitting algorithm

```
Non-splitted chromosome
/*Initialize relaxation coefficients CRv=vehicles, CRt=route duration, CRl=load*/
if the Current population= Pfeasiblethen then
    CRv←1 State CRt←1
    CRl←1
else
    CRv←crv /*Vehicles number relaxation coefficient */
    CRt←crt /*Maximum route duration relaxation coefficient */
    CRl←crl /*Maximum load relaxation coefficient */
end if
Q←Vehicle capacity × CRl
H←Maximum route duration ×CRt
Fc←Fleet per center × CRv
i←Number of centres
n←Number of nodes in the current period
Tour=(S1,S2,...,Sn) ←Giant tour of chromosome
$V_{1...n}$ ⟵ $\infty$ /*Initial costs of arcs */
$P_{1...n}$ ⟵ $\emptyset$ /*List of predecessors */
for v ⟵ 1 to n do
    /*Single node case*/
    load ⟵ Request Of ($S_v$)
    /*Add the node t to path*/
    path ⟵($S_v$)
    /*Choose the nearest centre dc to node v */
    /*If v corresponds to a VIP customer, choose its original centre */
    dc ⟵ Nearest Center ($S_v$)
    route ⟵ (dc, $S_v$, dc)
    time ⟵ Duration Of(route)
    s ⟵ v+1
end for
while s$\leq$ n and time¡H and load+RequestOf ($S_s$) $\leq Q$ do
    load⟵ load+Request Of ($S_s$)
    /* Add the nodes to the path */
    Add Node(path, Ss)
    /* Choose the centres that have enough vehicles (used vehicles ¡ Fc) */
    /* Compute, for each centre with s in the best placement, the route duration */
    /* Choose the centre dc with the minimal route duration */
    dc⟵ Best Center For (path)
    route ⟵ (dc, path, dc)
    time ⟵ Duration Of (route)
    if time$\leq$ $V_s$ then
        Vs ⟵ time
        $P_s$ ⟵ s-1
        s ⟵ s+1
    end if
end while
/* Using route and the list P to update the partial solutions */
Return Splitted chromosome
```

operation until reaching the parametrized number of nodes $N_{initial}$.These nodes are used to determine the first slope of the line connecting dc_i to the barycentre z of $(dc_i, n_{1,2} ..., n_{N_i nitial})$.

- Determine the slope $\Delta_c = \frac{y_2 - y_1}{x_2 - x_1}$ of the line connecting the center dc_i to z where (y_1, y_2) are the coordinates of dc_i and (x_1, x_2) are the coordinates of z.
- Add the node p which minimizes the value of D. Given the new slope Δ_t and the length l_p of the arc $N_{(initial,p)}$: $D = w_f |(\Delta_t - \Delta_c)/\Delta_c| + w_d \times l_p$. Where w_f= *shape weight*, w_d= *distance weight*, with $|\frac{\Delta_t - \Delta_c)}{\Delta_c}| \leq Max_{deviation}$
- Add the following nodes according to the same principle while respecting the non-relaxed constraints. If the deviation of the resulting slope Δ_c engendered by adding the node p' to the route is greater than $Max_{deviation}$, then the node p' will not be considered.
- Place the center dc_i as the last node of the route.
- Merge the routes of not fully loaded vehicles if the non-relaxed constraints allow it.

c) Step3: Route sequencing:

- Project the nodes coordinates of each route on a plane and divide it into two sub-routes by a split middle line (SML).
- Project the nodes of the first sub-route on the SML and reorder them in ascending order according to the resulting coordinates as shown in (see Fig. 2(a)).
- Project the nodes of the second sub-route on the SML and reorder them in descending order according to the resulting coordinates as shown in (see Fig. 2(a)).
- Reconstruct the route by connecting the two sub-routes considering dc_i as the starting and ending node of the route (see Fig. 2(a)).

d) Step4: Initial improvement:

- Perform inter-route improvement using a $swap(n1, n2)$ move by exchanging $n1$ successive customers of one route with $n2$ successive customers of another route while respecting the route's constraints.
- Perform intra-route improvement using an iterative $swap(1, 1)$ move.

The three populations $p_{feasible}$, p_{relax1} and p_{relax2} are filled using the intermediate solutions generated in Step 4.

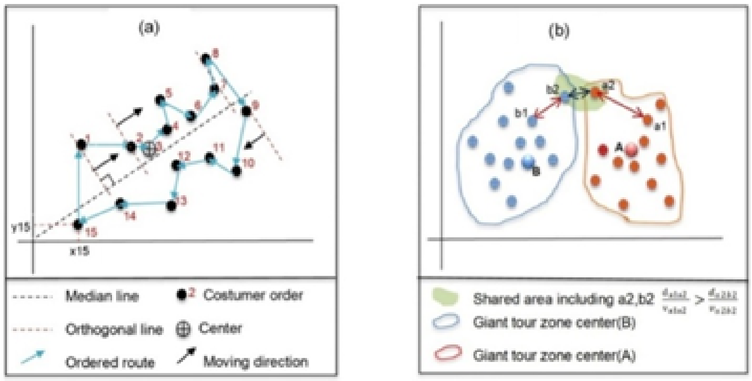


Fig. 2. (a) Route splitting and (b) The initial clustering

D. Selection and Crossover Mutation

In the selection procedure, the first parent is randomly selected from the first half of the population which is always sorted according to the fitness value, and the second parent is randomly chosen from the other half of the population. The proposed genetic operator performs three types of transpositions with different frequencies. To perform the crossover, two cutting points are defined randomly as shown in Fig. 3:

- A typical crossover consists of placing the genes between the cutting point of the second parent in the same position in the first offspring and the rest of its genes are copied from the first parent circularly. The same operation is performed for the second offspring by changing the roles of parents 1 and 2 as shown in (see Fig. 3).

- A transposition similar to the first one, but in this case the exchanged portions are partially or fully rotated (see Fig. 4).
- A transposition with position shifting in which the position of the exchanged portions is shifted when placed in the offspring chromosomes (see Fig. 5).

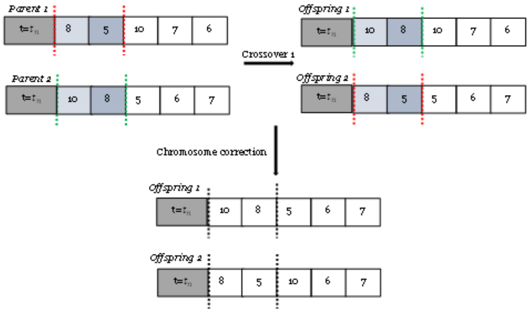


Fig. 3. First transposition method of the genetic operator.

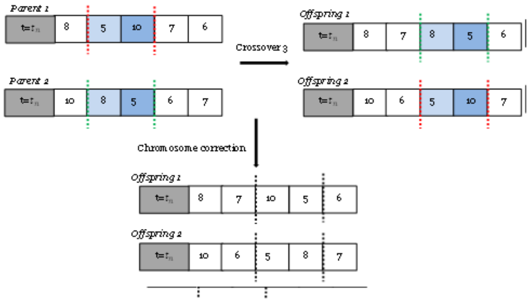


Fig. 4. Second transposition method of the genetic operator.

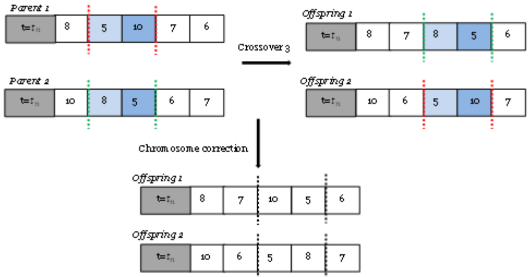


Fig. 5. Third transposition method of the genetic operator.

E. Local Variable Neighbourhood Tabu Search

The proposed VNS-Tabu algorithm (VNS-T) is a modified version of the Skewed Variable Neighbourhood Search algorithm proposed by Hansen et al. in 2020 [36], [37], it integrates a Tabu list and a variable acceptance margin α. The VNS-Tabu algorithm uses a list of neighbourhoods $k = (k_1, k_2, \ldots, k_n)$ corresponding to a sequence of intra-route and inter-route swap moves and shift moves applied on a current solution s. The local search (*LocalImprovement*) performs 2-opt and 3-opt heuristics moves. A new solution $S_i^{'}$ is accepted only if the cost quotient $\frac{f(s_i^{'})}{f(s)}$ is inferior to $1+\alpha$ where $\alpha = \beta \times \frac{i}{Maxreps}$ and β is a scale parameter. To avoid a local optimum, a *Tabulist* that stores the solution s and the move k_j performed on it is used. This list is emptied for each new value of α. The pseudo-code of **VNS-T** is presented in Algorithm 3.

Algorithm 3 : VNS-Tabu algorithm

```
$S_{best} \leftarrow$ Splitted chromosome
s$\leftarrow S_{best}$
$k = (k_1, k_2, ..., k_n) \leftarrow$ Set of neighbourhoods
j$\leftarrow$ 1
I$\leftarrow$1
$\alpha \leftarrow$0
while $i \leq Maxreps$ do
    while $j \leq n$ do
        if $(s, k_j) \notin Tabulist$ then
            $s^{'} \leftarrow$ Shake(s,$k_j$)
            $s^{''} \leftarrow LocalImprovements^{'}$
            if $\frac{f(s")}{f(s)} \leq (1+\alpha)$ then
                Tabulist.Add(s,$k_j$)
            end if
            if f(s")$\leq$ f(s) then
                $S_{best} \rightarrow s"$
                $s \leftarrow s''$
            else
                $j \leftarrow j+1$
            end if
        else$j \leftarrow j+1$
        end if
    end while
    Tabulist.Empty()
    $i \leftarrow i+1$
    $\alpha \leftarrow \beta \times \frac{1}{Maxreps}$
end while
Return
```

a) No results found by CPLEX within the time limit. b) The time limit for the small and medium sized instances was set to 3600s. For the large-scale instances (≥150 customers) the time limit is 7200s.

F. Results of the MPMA for the MDVRP Instances

To measure the efficiency of the MPMA, a comparative study is performed between the results of Vidal's HGA [31] and Juan's ILS [12] and those obtained by our algorithm with the best-known solutions (BKS) in the literature for 20 MDVRP instances of Cordeaux & al. Table V shows that the solutions obtained by the MPMA are close to the BKS with an average gap of 0.33%. Also, we observe that our method outperforms the ILS algorithm. Furthermore, MPMA was able to match 13 BKS within relatively small processing times. These experiments prove the efficiency of the proposed algorithm and its adaptability to solve the MDVR problem.

G. Fitness Evaluation

In the second echelon, the fitness evaluation for a given period t and a given population is based on the objective sub-function $T_3^t = \sum_{i,j \in I \cup J_t} \sum_{k \in K} \frac{d_{ij} \times x_{ij}^{kt} \times u_j^t \times \rho \times h_k}{100}$. For $p_{feasible}$, the fitness is equal to T_3^t. For p_{relax1} and p_{relax2}, the fitness is equal to $\frac{T_3^t}{Cf_v} \times Cf_l \times Cf_t$ where Cf_v is the percentage of centres respecting the fleet constraint,Cf_t is the percentage of routes respecting the duration constraint and Cf_l is the percentage of routes respecting the load constraint. For the first echelon, we use the objective function T_{total}.

TABLE IV. Comparison between the MILP and MPMA Results

Instances				MILP					MPMA			
Instance	Cst.	DCs.	Prds.	BI(1)	Gap(1-2)	LB(2)	Time(s)b	Stat.	Avg.(3)	Time(s)	Gap(3-2)(%)	Gap(3-1)(%)
C-PVip-1	20	4	2	10619	10.04	9553	3600	Int	9591	5	0.4	-10.72
C-PVip-2	20	4	2	10842	0.03	10838	336	Opt	10841	1	0.03	-0.01
C-PVip-3	30	4	2	11053	12.49	9673	3600	Int	9713	3	0.41	-13.8
C-PVip-4	30	4	2	11317	2.19	11069	3600	Int	11233	3	1.46	-0.75
C-PVip-5	30	2	3	8886	0.04	8882	3600	Int	8886	2	0.05	0
C-PVip-6	50	2	3	10746	8.64	9818	3600	Int	10097	3	2.76	-6.43
C-PVip-7	75	5	3	20224	4.28	19359	3600	Int	19451	17	0.47	-3.97
C-PVip-8	80	2	3	27384	11.19	24320	3600	Int	25146	9	3.28	-8.9
C-PVip-9	95	3	3	58096	2.11	56871	3600	Int	57711	15	1.46	-0.67
C-PVip-10	97	2	3	-a	-	12465	3600	-	13281	10	6.14	-
C-PVip-11	100	2	3	14604	23.38	11189	3600	Int	11510	17	2.79	-26.88
C-PVip-12	100	4	4	61560	2.31	60138	3600	OM	60661	10	0.86	-1.48
C-PVip-13	100	4	4	60207	5.55	56868	3600	Int	56952	47	0.15	-5.72
C-PVip-14	150	4	4	76222	14.91	64859	5405	OM	67327	13	3.67	-13.21
C-PVip-15	150	2	3	-	-	32355	7200	-	33909	13	4.58	-
C-PVip-16	200	4	4	-	-	69175	7200	-	71423	54	3.15	-
C-PVip-17	249	4	4	-	-	73482	2100	OM	79206	49	7.23	-
C-PVip-18	249	2	3	-	-	-	-	OM	50989	54	-	-
C-PVip-19	249	3	3	-	-	-	-	OM	31918	57	-	-

H. First Echelon Optimization using the Local Variable Neighbourhood Tabu Search Algorithm

The first phase of the MPMA generates the second echelon routes for each period t, then based on the best second echelon feasible solution S_{best}, the demands of the reassigned customers are calculated. These demands represent the quantities of goods to be delivered by the LC to each DC each period t. Additionally, the algorithm optimizes the semi-trailer route by performing the above-mentioned VNS-T algorithm.

V. Numerical Results

In this section, we present the results obtained by MPMA by solving 19 new instances built for the 2E-CMCPVRP-VCIS with IBM Ilog CPLEX Optimization Studio 20.1. After that, we compare the efficiency of our MPMA with Vidal's HGA and Juan's ILS based on Cordeaux's MDVRP benchmark instances. Three different scenarios are considered: non-collaborative, collaborative with VIP customers, and collaborative without VIP customers.

A. Description of Data Instances

19 new instances of different sizes and complexity are considered by adapting Cordeaux's MDVRP benchmark instances. For each instance, we add the following information related to the first and second echelon: semi-trailer's fuel consumption and annual maintenance costs, vehicles assignment to the centers, maintenance, annual leasing costs of each vehicle, centers' fixed and variable costs, the variable cost coefficient, customers' initial assignment to the centers, centers' delivery schedules, centers' VIP customers and third-party service costs. In these instances, the number of customers (Cst.) varies between 20 and 498, the number of centers (Dc) ranges from 2 to 6, the number of periods varies between 2 and 4, and the average speed of the arcs ranges between 70km/h and 110km/h. Moreover, the input parameter settings are: $H_{se} = 98l/100$km, H_{sl}=410l/100km,$h_(k)$ ∈[19,25]l/100km, F_v ∈[123 800,138000],L_v ∈[216000,288000], K_a = 190720km, M = 380000, B =2000, $\rho = 9.8$, $G_i \in [586, 645]$, P_i ∈[605,935], $y_i = 1.5$.

B. Parameter tuning

To tune the parameters of the proposed algorithm, multiple iterative computational treatments are performed on a set of MDVRP benchmark instances following an experimental methodology inspired by the design of experiments (DOE) approach. Firstly, we determine the three parameters that most significantly impact the performance of our MPMA and their levels through extensive testing. The parameters that obtained, and their respective levels are: $Maxreps \in \{10, 20, 40\}$, $Freq_{loc} \in \{2, 7, 14\}$, $P_{feasible} \in \{30, 50, 100\}$. In the second phase, 10 runs of the algorithm are performed on the instances set for each of the following twenty-seven parameter configurations: $\{10, 20, 40\} \times \{2, 7, 14\} \times \{30, 50, 100\}$. Then, the average of all the iterations' results for each instance is compared to the corresponding BKS. Next; the average gap and the standard deviation for each configuration is computed. The experiments indicate that the optimal parameter settings are: $maxreps = 40, freq_{loc} = 7, P_{feasible} = 30$.

C. Results of the MILP and MPMA for the 2E-CMCPVRP-VCIS Instances

The comparison between the MILP and MPMA results based on the values of the fitness function T_{total} are presented in Table IV. In the first four columns, the parameters of the 2E-CMCPVRP-VCIS instances are described in the following order: name of the instance, number of customers, number of centers, and number of periods. The second part of Table IV, from column 5 to column 9, summarizes the MILP results obtained by the CPLEX solver. BI refers to the best integer solution found by CPLEX, LB refers to the lower bound and column Stat describes the CPLEX status (Opt: CPLEX found

TABLE V. MPMA RESULTS VS BEST KNOWNS SOLUTIONS OF THE MDVRP

Instance	Cst.	DCs	Vehicles	BKS{1}	HGA{2}	ILS{3}	MPMA{4}	Gap(%) $\{4-1\}$	Gap(%) $\{4-2\}$	Gap(%) $\{4-3\}$	Time(s) (MPMA)
1	50	4	4	576.87	576.87	576.87	**576.87**	0	0	0	5
2	50	4	2	473.53	473.53	473.87	**473.53**	0	0	-0.07	3
3	75	5	3	641.19	641.19	641.19	**641.19**	0	0	0	5
4	100	2	8	1001.04	1001.04	1003.45	1003.86	0.28	0.28	0.04	43
5	100	2	5	750.03	750.03	751.9	**750.03**	0	0	-0.25	16
6	100	3	6	876.5	876.5	876.5	**876.5**	0	0	0	16
7	100	4	4	881.97	881.97	885.19	**881.97**	0	0	-0.36	50
8	249	2	14	4372.78	4372.78	4409.23	4400.36	0.63	0.63	-0.2	273
9	249	3	12	3858.66	3858.66	3882.58	3882.11	0.61	0.61	-0.01	296
10	249	4	8	3629.6	3631.11	3646.67	3634.74	0.14	0.1	-0.33	257
11	249	5	6	3545.48	3546.06	3547.09	3546.06	0.02	0	-0.03	426
12	80	2	5	1318.95	1318.95	1318.95	**1318.95**	0	0	0	25
13	80	2	5	1318.95	1318.95	1318.95	**1318.95**	0	0	0	9
14	80	2	5	1360.12	1360.12	1360.12	**1360.12**	0	0	0	25
15	160	4	5	2505.42	2505.42	2511.92	**2505.42**	0	0	-0.26	78
16	160	4	5	2572.23	2572.23	2573.78	**2572.23**	0	0	-0.06	58
17	160	4	5	2709.09	2709.09	2709.09	**2709.09**	0	0	0	17
18	240	6	5	3702.85	3702.85	3702.85	3708.7	0.16	0.16	0.16	364
19	240	6	5	3827.06	3827.06	3840.91	**3827.06**	0	0	-0.36	213
20	240	6	5	4058.07	4058.07	4063.64	4091.78	0.83	0.83	0.69	278

the optimal solution, OM: CPLEX goes out of memory, Int: The best solution was found by CPLEX within the time limit). The results obtained show that CPLEX was able to find an optimal solution only for the small-sized instance C-PVip-2 within 336 seconds. For sixteen instances, CPLEX was able to find a feasible solution within the set time limit and with an average gap between BI and LB of 7.7%. For two instances, the solver found a feasible solution but went out of memory before the time limit. For the large-scale instances, CPLEX went out of memory before finding any solution or LB. The third-party of Table IV, from columns 10 to 13, presents the average results of the MPMA on 10 runs for each one of the 19 instances, the computational time of the MPMA and the gaps between the MPMA results and the LB and BI respectively. The results revealed that the MPMA was able to find good quality solutions for all the instances with an average gap of 2.01% with the LB and -6.59% with the BI and within a relatively short average computational time of 43.23 seconds. Moreover, the gap between the MPMA and CPLEX for the instance C-PVip-2 for which the MILP optimal solution was found is negligible. The computational results validate the MILP model and prove the efficiency of the MPMA.

VI. CONCLUSION

In this paper, we have introduced the two-echelon collaborative multi-center multi-periodic vehicle routing problem with VIP customers and inflexible delivery schedules (2E-CMCPVRP-VCIS). To solve the proposed model, an efficient multi-phase solving approach (MPSA) based on a multi-population memetic algorithm (MPMA) and a variable neighborhood search method is proposed. The performance of the MPSA is evaluated on benchmark MDVRP instances as well as newly created instances for our case of study. The numerical results on the benchmark instances show that the MPMA outperforms Juan's ILS and can find high-quality solutions with an average gap of 0.33% to the BKS. Furthermore, the results on the new instances prove the performance of our algorithm and its superiority compared to the upper bounds obtained by solving the MILP model on CPLEX. This research underscores the effectiveness of the proposed MPSA in tackling the 2E-CMCPVRP-VCIS.

REFERENCES

[1] R. G. Thompson & P.K. Hassall, "A Collaborative Urban Distribution Network." Procedia - Social and Behavioral Sciences 39: 230–40. https://doi.org/10.1016/j.sbspro.2012.03.104, 2012.

[2] C. Laisney, "Covid-19 et comportements alimentaires." Futuribles N°437 (4): 83. https://doi.org/10.3917/futur.437.0083, 2020.

[3] M. Niels, L. Verdonck, A. Caris & B. Depaire, "Horizontal Collaboration in Logistics: Decision Framework and Typology." Operations Management Research 11 (1–2): 32–50. https://doi.org/10.1007/s12063-018-0131-1, 2018.

[4] J. Gonzalez-Feliu & J.M. Salanova, "Defining and Evaluating Collaborative Urban Freight Transportation Systems." Procedia - Social and Behavioral Sciences 39: 172–83. https://doi.org/10.1016/j.sbspro.2012.03.099, 2012.

[5] M. Savelsbergh & T.V Woensel, "50th Anniversary Invited Article—City Logistics: Challenges and Opportunities." Transportation Science 50 (2): 579–90. https://doi.org/10.1287/trsc.2016.0675, 2016.

[6] Y. Wang, M. Xiaolei, Z. Li,Y. Liu, M. Xu & Y. Wang. "Profit Distribution in Collaborative Multiple Centers Vehicle Routing Problem." Journal of Cleaner Production 144 (February): 203–19. https://doi.org/10.1016/j.jclepro.2017.01.001, 2017.

[7] B. Dai & H. Chen, "Mathematical Model and Solution Approach for Carriers' Collaborative Transportation Planning in Less than Truckload Transportation." International Journal of Advanced Operations Management 4 (1/2): 62. https://doi.org/10.1504/IJAOM.2012.045891, 2012.

[8] H. Nagati, C. Rebolledo & M. Jobin, "Collaboration entre les acteurs de la chaîne logistique: conditions de succés." Gestion 34 (1): 27. https://doi.org/10.3917/riges.341.0027, 2009.

[9] J. Gonzalez-Feliu & J. Morana, "Collaborative Transportation Sharing: From Theory to Practice via a Case Study from France," 23, 2011.

[10] Y. Wang, J. Zhang, K. Assogba, Y. Liu, M. Xu & Y. Wang, ”collaboration and Transportation Resource Sharing in Multiple Centers Vehicle Routing Optimization with Delivery and Pickup.” Knowledge-Based Systems 160 (November): 296–310. https://doi.org/10.1016/j.knosys.2018.07.024, 2018.

[11] T. Vidal, G.C. Teodor, M. Gendreau, N. Lahrichi & W. Rei, “A Hybrid Genetic Algorithm for Multidepot and Periodic Vehicle Routing Problems.” Operations Research 60 (3): 611–24. https://doi.org/10.1287/opre.1120.1048, 2012.

[12] A. Rahimi-vahed, G.G Crainic, M. Gendreau & W. Rei, “Fleet-Sizing for Multi-Depot and Periodic Vehicle Routing Problems Using a Modular Heuristic Algorithm.” Computers & Operations Research 53 (January): 9–23. https://doi.org/10.1016/j.cor.2014.07.004, 2015.

[13] T. Ramos, I.G Marcio & A.P.B Póvoa, ”Multi-Depot Vehicle Routing Problem: A Comparative Study of Alternative Formulations.” International Journal of Logistics Research and Applications 23: 103-120. doi:10.1080/13675567.2019.1630374,2020.

[14] R. Sharma & S. Saini, “Heuristics and Meta-Heuristics Based Multiple Depot Vehicle Routing Problem: A Review.” In International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 683-689. doi: 10.1109/ICESC48915.2020.9155814, 2020.

[15] C.K.Y Lin, “A Cooperative Strategy for a Vehicle Routing Problem with Pickup and Delivery Time Windows.” Computers & Industrial Engineering 55 (4): 766–82. https://doi.org/10.1016/j.cie.2008.03.001, 2008.

[16] M. Torres, R. Jairo, A. Muñoz-Villamizar, A. Carlos & V. Mejía, “On the Impact of Collaborative Strategies for Goods Delivery in City Logistics.” Production Planning & Control 27 (6): 443–55. https://doi.org/10.1080/09537287.2016.1147092, 2016.

[17] M. Apichit & A. Udomsakdigool, “Solving the Collaborative Bidirectional Multi-Period Vehicle Routing Problems under a Profit-Sharing Agreement Using a Covering Model.” International Journal of Industrial Engineering Computations, 185–200. https://doi.org/10.5267/j.ijiec.2019.10.002, 2020.

[18] M. Guajardo, “Environmental Benefits of Collaboration and Allocation of Emissions in Road Freight Transportation.” In Sustainable Freight Transport, edited by Vasileios Zeimpekis, Emel, 2018.

[19] C. Catherine, C. Cottrill, J.F Ehmke & K. Tierney, “Collaborative Urban Transportation: Recent Advances in Theory and Practice.” European Journal of Operational Research 273 (3): 801–16. https://doi.org/10.1016/j.ejor.2018.04.037, 2019.

[20] X. Wang, H. Kopfer & M. Gendreau, “Operational Transportation Planning of Freight Forwarding Companies in Horizontal Coalitions.” European Journal of Operational Research 237 (3): 1133–41. https://doi.org/10.1016/j.ejor.2014.02.056, 2015.

[21] M. Soysal, M. Jacqueline, B. Bloemhof-Ruwaard, R. Haijema & G.A.J Jack, “Modeling a Green Inventory Routing Problem for Perishable Products with Horizontal Collaboration.” Computers & Operations Research 89 (February): 168–82. https://doi.org/10.1016/j.cor.2016.02.003, 2016.

[22] P. Cuervo, D.C Vanovermeire & K. Sörensen, “Determining Collaborative Profits in Coalitions Formed by Two Partners with Varying Characteristics.” Transportation Research Part C: Emerging Technologies 70 (September): 171–84. https://doi.org/10.1016/j.trc.2015.12.011, 2016.

[23] H. Vera, K. Smilowitz & L. de la Torre, “A Periodic Location Routing Problem for Collaborative Recycling.” IISE Transactions 49 (4): 414–28. https://doi.org/10.1080/24725854.2016.1267882, 2017.

[24] Z. Lin, R. Baldacci, D. Vigo & X. Wang, ”Multi-Depot Two-Echelon Vehicle Routing Problem with Delivery Options Arising in the Last Mile Distribution.” European Journal of Operational Research 265 (2): 765–78. https://doi.org/10.1016/j.ejor.2017.08.011, 2018.

[25] Y. Wang, S. Zhang, K. Assogba, J. Fan, M. Xu & Y. Wang,”Economic and environmental evaluations in the two-echelon collaborative multiple centers vehicle routing optimization”, Journal of Cleaner Production Volume 197, Part 1, Pages 443-461, ISSN 0959-6526, https://doi.org/10.1016/j.jclepro.2018.06.208, 2018.

[26] Y. Wang, Y. YUAN, G. Xiangyang, X. Maozeng, W. Li, W. Haizhong & Y. Liu, ”Collaborative two-echelon multicenter vehicle routing optimization based on state–space–time network representation”. Journal of Cleaner Production. 258. 120590. 10.1016/j.jclepro.2020.120590, (2020).

[27] C. Cleophas, C. Cottrill, J.F. Ehmke & K. Tierney, ”Collaborative urban transportation: Recent advances in theory and practice”, European Journal of Operational Research, Volume 273, Issue 3, 2019,Pages 801-816, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2018.04.037.

[28] J. Li, G. Rong & Y. Feng, “Request Selection and Exchange Approach for Carrier Collaboration Based on Auction of a Single Request.” Transportation Research Part E: Logistics and Transportation Review 84 (December): 23–39. https://doi.org/10.1016/j.tre.2015.09.010, 2015.

[29] X.S. Xiu, Q. G. Huang & M. Cheng, “Truthful, Budget-Balanced Bundle Double Auctions for Carrier Collaboration.” Transportation Science 51 (4): 1365–86. https://doi.org/10.1287/trsc.2016.0694, 2017.

[30] M. Gansterer, F.R Hartl & M. Savelsbergh, “The Value of Information in Auction-Based Carrier Collaborations.” International Journal of Production Economics 221 (March): 107485. https://doi.org/10.1016/j.ijpe.2019.09.006, 2020.

[31] A. Alp, N. Agatz, L. Kroon & R. Zuidwijk, “Crowdsourced Delivery—A Dynamic Pickup and Delivery Problem with Ad Hoc Drivers.” Transportation Science 53 (1): 222–35. https://doi.org/10.1287/trsc.2017.0803, 2019.

[32] L. Quintero-Araujo, A. Carlos, A. Gruler & J. Faulin, “Using Horizontal Cooperation Concepts in Integrated Routing and Facility-Location Decisions.” International Transactions in Operational Research 26 (2): 551–76. https://doi.org/10.1111/itor.12479, 2019.

[33] W. Ho, T.S George, P. Ji & C.W Lau, “A Hybrid Genetic Algorithm for the Multi-Depot Vehicle Routing Problem.” Engineering Applications of Artificial Intelligence 21 (4): 548–57. https://doi.org/10.1016/j.engappai.2007.06.001, 2008.

[34] A. El-Fallahi, C. Prins & R. W. Calvo,”A memetic algorithm and a tabu search for the multi-compartment vehicle routing problem”, Computers & Operations Research, Volume 35, Issue 5, 2008, Pages 1725-1741, ISSN 0305-0548,https://doi.org/10.1016/j.cor.2006.10.006.

[35] Y. Shi, L.V Lingling, F. Hu & Q. Han, “A Heuristic Solution Method for Multi-Depot Vehicle Routing-Based Waste Collection Problems.” Applied Sciences 10 (7): 2403. https://doi.org/10.3390/app10072403, 2020.

[36] Vidal, T. 2014. “Implicit Depot Assignments and Rotations in Vehicle Routing Heuristics.” European Journal of Operational Research, 14.

[37] P. Hansen, N. Mladenović & J.A. Moreno-Pérez, ”Variable neighbourhood search: Methods and applications.” Annals of Operations Research 175: 367-407. doi: 10.1007/s10479-009-0657-6, 2010.

Speech Recognition Models for Holy Quran Recitation Based on Modern Approaches and Tajweed Rules: A Comprehensive Overview

Sumayya Al-Fadhli[1], Hajar Al-Harbi[2], Asma Cherif[3]
Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia[1,2]
Department of Computer Science-Adham University College, Umm Al-Qura University, Makkah, Saudi Arabia[1]
Department of Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia[3]
Center of Excellent in Smart Environment Research, King Abdulaziz University, Jeddah, Saudi Arabia[3]

***Abstract*—Speech is considered the most natural way to communicate with people. The purpose of speech recognition technology is to allow machines to recognize and understand human speech, enabling them to take action based on the spoken words. Speech recognition is especially useful in educational fields, as it can provide powerful automatic correction for language learning purposes. In the context of learning the Quran, it is essential for every Muslim to recite it correctly. Traditionally, this involves an expert *gari* who listens to the student's recitation, identifies any mistakes, and provides appropriate corrections. While effective, this method is time-consuming. To address this challenge, apps that help students fix their recitation of the Holy Quran are becoming increasingly popular. However, these apps require a robust and error-free speech recognition model. While recent advancements in speech recognition have produced highly accurate results for written and spoken Arabic and non-Arabic speech recognition, the field of Holy Quran speech recognition is still in its early stages. Therefore, this paper aims to provide a comprehensive literature review of the existing research in the field of Holy Quran speech recognition. Its goal is to identify the limitations of current works, determine future research directions, and highlight important research in the fields of spoken and written languages.**

***Keywords*—Speech recognition; acoustic models; language model; neural network; deep learning; quran recitation**

I. Introduction

Speech is the most natural way to communicate with people [7]. Designing a machine that mimics human behavior, including speaking naturally and responding correctly to spoken language, has puzzled engineers and scientists for centuries [51]. Automatic speech recognition (ASR) refers to the computational process of transforming acoustic speech signals into written words or other linguistic units through dedicated algorithms [28], [7]. The goal of ASR is to enable machines to interpret and respond to spoken language [4]. ASR involves the capability of a machine to accurately recognize speech, convert it into text, and take appropriate actions based on human instructions [7]. In particular, speech recognition is useful in educational fields as it allows for the building of powerful automatic correctors for language learning purposes. As [41], they build a model of English pronunciation learning for Chinese learners.

The researchers have made significant contributions to speech processing in various languages spoken worldwide. There are three classes in Arabic, which has approximately 420 million speakers [47]. The primary class taught in schools is Modern Standard Arabic (MSA), which adheres to the grammatical rules of the Arabic language. The second class is Arabic Dialect (AD), which represents the everyday spoken language of native Arabic speakers, varying across countries and regions. The third class is classical Arabic (CA), the language used in the Holy Quran, which has been renowned globally for centuries. CA is known for its extensive grammar and vocabulary, as well as its unique recitation guidelines [15], [24].

Recently, the use of speech recognition in the Quranic recitation field has emerged as an important research direction. Indeed, there are more than two billion Muslims in the world [2]. Muslims generally strive to learn the precise recitation of CA and adhere to certain rules known as *Tajweed* in order to recite the Holy Quran accurately. Learning these rules is very important for all Muslims to master the recitation of the Holy Quran [15]. Consequently, building accurate Holy Quran Speech Recognition (HQSR) models represents a significant research outcome for all Muslims.

Teaching the correct recitation of the Quran is essential for every Muslim. Learning Quran recitation usually depends on an expert, also known as *gari*, who listens to the student's recitation, determines recitation mistakes, and instructs the student with the appropriate correction. This way of learning is very effective, but it's time-consuming because the teacher needs to correct the errors of every student independently. For this reason, the apps that help students fix their recitation of the Holy Quran are beneficial and essential, but these apps need a robust and error-free speech recognition model. Despite conducting several research studies in this area, researchers have not yet achieved the optimal solution for recognizing speech in the Holy Quran. Though recent models have been applied to written and spoken Arabic and non-Arabic speech recognition and produced highly accurate results, Quran speech recognition is still in its early stages. Therefore, this paper aims to propose a comprehensive literature review of the works in the field of Holy Quran speech recognition and shed light on some important research in the field of spoken and written languages.

The main motivation for our research is as follows:

1) Though the Holy Quran represents an essential book for all Muslims, current models for Holy Quran speech recognition have low accuracy or do not cover all chapters (i.e., rely on small datasets).
2) Some people find it difficult to attend Quran learning courses or retrieve their memorization in front of the teacher. Many individuals struggle to retain the Quran due to fear. Thus, building a professional app for Quran learning is important to help them retain the Quran in their home.
3) Quran memorization requires a continuous review process, which is time-consuming. Thus, it is hard for Quran teachers to listen and validate long recitations for many students.
4) Some people prefer reading what they memorize, especially in night prayer (i.e., without reading from the *Mushaf* to not lose their submission in prayer). However, they can easily make mistakes. An automatic corrector can assist Muslims in their prayers.
5) Some non-Arabic countries, mainly those with a minority of Muslims, do not have enough qualified teachers to teach the Holy Quran.

However, research in the field of HQSR is still in its early stages. Indeed, recognizing individual words is easy, but the challenge is recognizing continuous recitation [7] and detecting erroneous recitation and violations of tajweed rules. In the realm of speech recognition systems (see Fig. 1), various factors, such as speaker dependency, vocabulary size, and noisy environments, can significantly impact their performance. Recognition performance increases with limited vocabulary and reciter-dependent conditions while using broad vocabulary and reciter-independent scenarios; performance can decrease significantly [7]. Besides, most research developments focus on one or a few chapters or a few tajweed rules. Also, existing works in HQSR suffer from the lack of large datasets used. Finally, current works use traditional techniques and do not investigate end-to-end learning.

The critical objective of this research is to use machine learning for Holy Quran recitation. Our main contributions are to provide a thorough literature review to find the most important issues that need more investigation in the field of Holy Quran speech recognition. Moreover, our study summarizes some important and informative papers in the Arabic and non-Arabic languages fields and recent papers in the HQSR field and provides a taxonomy for speech recognition and HQSR.

Various machine learning algorithms could be used in speech recognition, including Dynamic Time Warping (DTW), Hidden Markov Models (HMM), and Artificial Neural Networks (ANN) [51].

In the context of Arabic ASR, many algorithms were used, such as recurrent neural networks (RNN), long short-term memory (LSTM), which is a particular case of RNN, and connectionist temporal classification (CTC) [4].

The remaining parts of this paper are structured in the following way: Section II discusses some important research in written and spoken languages speech recognition. Section III discusses the recent papers in the field of Holy Quran Speech Recognition. Next, Section IV discusses some of the research directions in the field of Holy Quran Speech Recognition. Finally, Section V concludes the paper.

II. Speech Recognition for Written & Spoken Languages

In this section, we highlight some speech recognition solutions that produced impressive results in Arabic and non-Arabic languages. These solutions may be categorized as traditional speech recognition (either with deep learning architecture or without deep learning) or as an end-to-end-based speech recognition solution.

A. Traditional Models

Fig. 2 shows that traditional ASR systems are made up of three separate parts: the acoustic model, the pronunciation model, and the language model [54]. The Acoustic Model (AM) assesses the likelihood of acoustic units such as phonemes, graphemes, or sub-word units [13]. In contrast, the Language Model (LM) evaluates the likelihood of word sequences. By integrating linguistic knowledge derived from extensive text collections, language models improve the precision of acoustic models. These models use the acquired syntactic and semantic rules to re-evaluate the hypotheses generated by the acoustic model. The process of mapping a series of phonemes to words is done by the Pronunciation Dictionary (PD), and it aligns the phonetic transcriptions produced by the AM system with the unprocessed text used in language models. The training of these three components is done individually, and then they are merged together to form a search graph by utilizing finite-state transducers (FSTs). Feature Extraction (FE) takes input speech as input, produces the essential features, and then sends these features to the decoder. Following that, the decoder produces lattices, which are then evaluated and ordered to generate the desired sequences of words.

The acoustic model can be modeled using HMMs [32] and Gaussian Mixture Models (GMMs) [61]. It is worth noting that recent ASR models have replaced the use of GMMs in the acoustic model with deep neural networks (DNNs) [30]. These are referred to as hybrid HMM-DNN and are widely used as competitive ASR models. Also, some research replaced GMMs with Bidirectional Long-Short Term Memory (BLSTM) [50], while some other studies replaced HMM with another classification method such as Support Vector Machine (SVM) [12], [36], Linear Discriminant Analysis (LDA) combined with Quadratic Discriminant Analysis (QDA) [33], Convolutional Neural Networks (CNNs) and SVM [40], and Hidden Semi-Markov Model (HSMM) [34], etc.

Some research has been suggested to address speech recognition for Arabic and non-Arabic languages.

Arabic Language. In their study, [39] introduced a novel approach that combines three distinct training systems for speech recognition. Four-gram language model re-scoring, system combination with minimum Bayes risk decoding, and lattice-free maximum mutual information are a few of these groups. They achieved significant progress, with a word error rate of 42.25% on the Multi-Genre Broadcast (MGB-3) Arabic

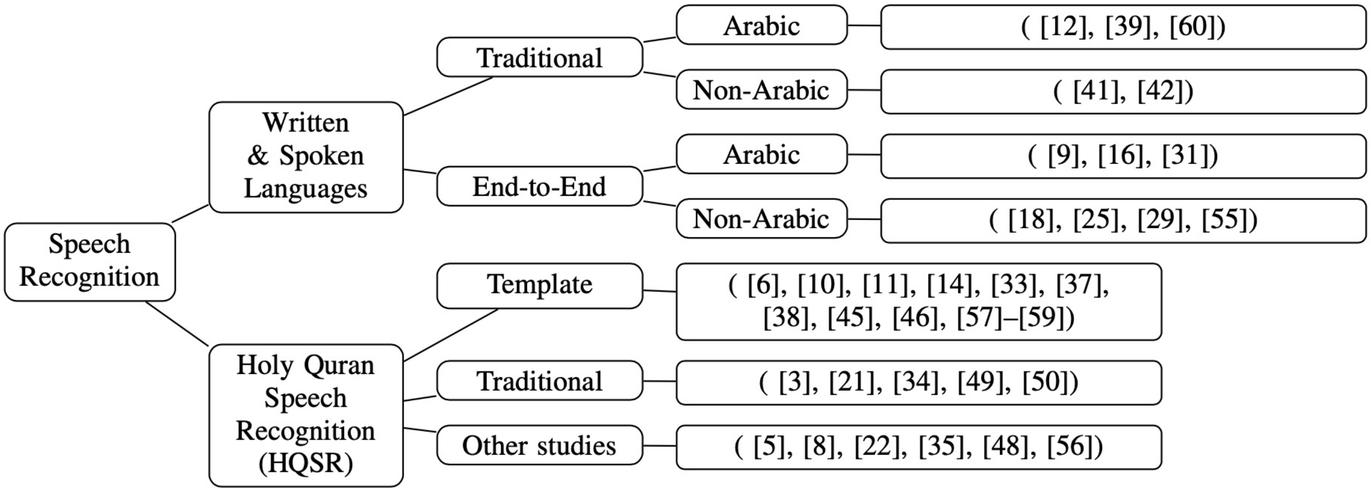


Fig. 1. Speech recognition taxonomy.

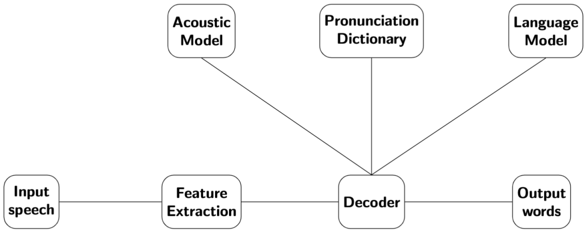


Fig. 2. Traditional ASR Pipeline ([54]).

development set. They got this result by using a 4-gram rescoring strategy for a chain BLSTM system. This system did better than a DNN system that had a word error rate of 65.44%.

In [60], the authors presented a comprehensive framework for Arabic speech recognition. To turn sequences of Mel Frequency Cepstral Coefficients (MFCC) and Filter Bank (FB) features into fixed-size vectors, they used recurrent LSTM or GRU architectures. They then fed these vectors into a multi-layer perceptron network (MLP) to perform classification and recognition tasks. The researchers evaluated their system using two different databases: one for spoken-digit recognition and another for spoken TV commands. However, a limitation of their work is the absence of datasets that incorporate recorded speech signals in noisy, realistic environments.

In [12], they presented a speech recognition system for the Arabic language. The system aimed to evaluate three feature extraction algorithms: MFCC, Power Normalized Cepstral Coefficients (PNCC), and Modified Group Delay Function (ModGDF). We performed the classification process using an SVM. The results indicated that PNCC was the most effective algorithm, while ModGDF achieved moderate accuracy. PNCC and ModGDF outperformed MFCC in terms of precision. PNCC achieved an accuracy rate of 93% to 97%, ModGDF achieved 90%, and MFCC achieved 88%.

Non-Arabic languages. The authors of [42] presented the design of Kaldi, a speech recognition toolkit that is freely available and open-source. The highly permissive Apache License v2.0, under which Kaldi is released, enables extensive usage. Kaldi provides a robust speech recognition system that utilizes finite-state transducers and is built on the OpenFst library. The toolkit provides comprehensive documentation and scripts that make it easier to build comprehensive recognition systems. Kaldi is coded in C++, and its core library offers a range of functionalities, including phonetic-context modeling, acoustic modeling using subspace Gaussian mixture models (SGMM), standard Gaussian mixture models, and linear and affine transforms.

A speech-learning system for the English language was developed and implemented by [41]. It utilized a speech recognition technique based on HMM to decode speech using the Viterbi algorithm and determine the recognition score through posterior probability. The system achieved an average recognition rate of 94%. Its purpose was to help English learners assess pronunciation accuracy during verbal practice and identify different types of errors. By engaging in systematic practice with this system, users can significantly enhance their listening and speaking skills. The system provides real-time feedback on oral pronunciation accuracy, error correction reports, and allows for repeated practice to facilitate effective training.

Table I summarizes traditional speech recognition techniques for Arabic and non-Arabic languages.

B. End-to-End-based Speech Recognition Models

The purpose of the end-to-end (E2E) system is to directly transform a series of acoustic features into a corresponding series of graphemes or words. This approach greatly simplifies traditional speech recognition methods by eliminating the need for manual labeling of information in the neural network. Instead, the E2E system automatically learns language and pronunciation information, as depicted in Fig. 3.

End-to-end speech recognition systems typically rely on an encoder-decoder framework. According to studies [17], [13], this architecture takes an audio file as input and processes it through a series of convolution layers to generate a condensed

TABLE I. Summary of Traditional Speech Recognition Techniques in Written and Spoken Languages

Ref	Lang.	Main idea	FE	Classification	LM	AM
[12]	Arabic	Study three feature extraction methods, MFCC, PNCC, and ModGDF for the development of an ASR System in Arabic.	MFCC, PNCC, and ModGDF	SVM	-	-
[60]	Arabic	Prsent an approach based on RNN to process sequences of variable lengths of MFCCs, FBs and delta-delta features of the different spoken digits/commands.	MFCC (static and dynamic features), and the FB coefficients.	LSTM and Neural Network (MultiLayer Perceptron: MLP) classifier	-	BiLSTM model
[39]	Arabic	Improve Hybrid ASR for MGB & Al-Jazeera speech data.	MFCCs features	Hybrid ASR using TDNN-LSTM & Bi-directional prioritized grid LSTM (BPGLSTM)	n-gram LM	Hybrid ASR using TDNN-LSTM & BPGLSTM
[42]	Non-Arabic	Build a new open source toolkit for Conventional speech recognition from scratch called Kaldi toolkit.	MFCC features	GMM-DNN-HMM	bigram	DNN-HMM
[41]	Non-Arabic	Build a leight-weight speech recognition using GMM-HMM, to learn English language using HMM based speech recognition for Chinese speakers.	MFCC features	HMM based	n-gram	GMM-HMM

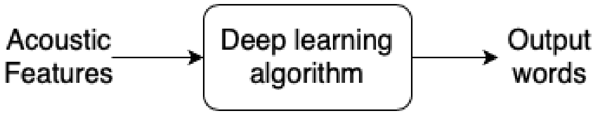


Fig. 3. End-to-End ASR Pipeline ([54]).

vector. The decoder then uses this vector to generate a character sequence. Researchers can use different objective functions, such as CTC [23], ASG [20], LF-MMI [25], sequence-to-sequence [18], transduction [44], and differentiable decoding [19], to optimize the end-to-end ASR [13]. Researchers have also explored different neural network architectures, including ResNet [27], TDS [26], and Transformer [52]. Additionally, integrating an external language model has been shown to improve the overall performance of the system.

Recently, some research has been suggested for both Arabic and non-Arabic speech recognition using end-to-end models. In what follows, we discuss and summarize these solutions.

Arabic Language. In [9], the researchers introduced the first comprehensive approach to building an Arabic speech-to-text transcription system. They utilized lexicon-free RNNs and the CTC objective function to achieve this. The system consisted of three main components: a BDRNN acoustic model, a language model, and a character-based decoder. Unlike word-level decoders, their decoder did not rely on a lexicon during the transcription process. The RNN acoustic and language model successfully distinguished between characters with the same accent but different writing styles. The researchers evaluated the model using a 1200-hour corpus of Aljazeera multi-genre broadcast programs, resulting in a 12.03% word error rate for non-overlapped speech. It is important to note that deep learning techniques were only used in the feature extraction phase.

In [16], the authors proposed a robust diacritized ASR system using both traditional ASR and end-to-end ASR techniques. They trained and tested their models on the Standard Arabic Single Speaker Corpus (SASSC) with diacritized text data, using MFCCs and FB for feature extraction. The ASR speech recognition system incorporated a total of eight models, comprising four GMM models, two SGMM models, and two DNN models. We constructed these models using the KALDI toolkit and performed language modeling using CMU-CLMTK. The best achieved word error rate (WER) among these models was 33.72% using DNN-MPE. Additionally, the authors proposed an end-to-end approach for diacritized Arabic ASR, employing joint CTC-attention and CNN-LSTM attention methods. The CNN-LSTM with attention method outperformed the others, achieving a character error rate (CER) of 5.66% and a WER of 28.48%. This method resulted in a significant reduction in WER compared to both the traditional ASR and joint CTC-attention method, by 5.24% and 2.62%, respectively.

Researchers conducted a comprehensive comparison on Arabic language and its dialects using different ASR approaches in [31]. The researchers collected a new evaluation set comprising news reports, conversational speech, and various datasets to ensure unbiased analysis. They extensively analyzed the errors and compared the ASR system's performance with that of expert linguists and native speakers. While the machine ASR system showed better performance than the native speaker, there was still an average WER gap of 3.5% compared to expert linguists in raw Arabic transcription. The proposed end-to-end transformer model outperformed prior state-of-the-art systems on MGB2, MGB3, and MGB5 datasets, achieving new state-of-the-art performances of 12.5%, 27.5%, and 33.8%, respectively.

Non-Arabic Languages. Researchers introduced ESPnet, a novel open-source platform for end-to-end speech processing, in [55]. ESPnet leverages dynamic neural network toolkits like Chainer and PyTorch, serving as the primary deep learning engine. This platform simplifies the training and recognition processes of the entire ASR pipeline. The ESPnet uses the same feature extraction/format, data processing, and scheme style as the Kaldi ASR toolkit. This gives researchers a complete way to test speech recognition and other speech processing techniques. The test results show that ESPnet does a good job with ASR and is about as efficient as the most advanced HMM/DNN systems that use traditional setups. Notably, ESPnet has made significant advancements, including the incorporation of multi-GPU functionality (up to 5 GPUs). In just 26 hours, ESPnet successfully completed the training of 581 hours of the CSJ task.

In [25], the authors described a simple HMM-based end-to-end method for ASR and tested how well it worked on well-known large-vocabulary speech recognition tasks, specifically the Switchboard and Wall Street Journal (WSJ) corpora. The authors trained the acoustic model used in this approach without the need for initial alignments, prior training, pre-estimation, or transition training, making it entirely neural except for the decoding/LM part. The proposed method surpassed other end-to-end methods in similar setups, particularly when dealing with small databases. By employing a comprehensive biphone modeling approach, the researchers achieved results almost comparable to regular LF-MMI training.

In [18], the researchers introduced a new attention-based model called Listen, Attend, and Spell (LAS) for sequence-to-sequence speech recognition. LAS combines the sound, pronunciation, and language model parts of regular ASR systems into a single neural network, so there's no need for a separate dictionary or text normalization. The researchers compared LAS with a hybrid HMM-LSTM system and found that LAS achieved a WER of 5.6%, outperforming the hybrid system's WER of 6.7%. In a dictation task, LAS achieved a WER of 4.1%, while the hybrid system achieved a WER of 5%.

The study [29] introduced a new strategy to address multilingual ASR speech recognition, specifically in the context of code-switching speech. The researchers employed three techniques to achieve this. The researchers decoded the speech by utilizing a global language model constructed from multilingual text. Their system used a multigraph approach along with weighted finite-state transducers (WFST), which let them switch between languages while decoding by using a closure operation. The output of this process was a bilingual or multilingual text based on the input audio. Secondly, they employed a robust transformer system for speech decoding. They found that WFST decoding was particularly suitable for inter-sentential code-switching datasets among the techniques used.

Table II summarizes end-to-end speech recognition techniques for Arabic and non-Arabic languages.

In the following, we compare the end-to-end architecture mentioned in the previous works with the baseline traditional techniques on the same datasets as indicated in the previous reviewed studies. As we see in Table III, the end-to-end architecture outperforms the hybrid architecture in all the studies mentioned in the table, except [25].

III. Holy Quran Speech Recognition (HQSR)

This section summarizes the recent studies that concern the HQSR. It also presents the used techniques and the performance of the proposed solutions. Moreover, it determines the gap and limitations in the current research on the HQSR. We classified the current studies of HQSR into three categories: template-based speech recognition, traditional-based speech recognition, and other HQSR studies.

A. Template Based Speech Recognition

This section summarizes the papers that follow template-based speech recognition, which is an old style of speech recognition that relied only on FE, classification, and matching techniques (i.e., it didn't have acoustic, lexical, or language models).

In [46], the authors provide a deep learning model utilizing a dataset of seven famous reciters and CNNs. They employ MFCCs to extract and assess data from audio sources. Their provided model achieved 99.66% accuracy.

In [6], the authors highlight the key distinctions between a basic ASR system and an ASR-based language tutor specifically designed for Quran memorization. They demonstrate that ASR techniques alone are not sufficient for an intelligent Quran tutor and propose modifications to enhance its capabilities. To support their claims, the researchers utilize data from Sūrat Al-Nass. However, one of the major obstacles to developing a Quran tutor is the absence of a comprehensive dataset containing both correct and incorrect recitations, which is necessary for conducting meaningful experiments in this domain.

In [37], researchers proposed an online verification system for Quran verses to ensure the integrity and authenticity of the Quran. They gathered data from ten expert Qari who recited Surat Al-Nass ten times correctly and ten times with various types of mistakes (e.g., Tajweed, Makhraj, missing words). Unlike modern techniques, this study did not utilize acoustic, lexical, or language models. Instead, it relied on MFCC for feature extraction and HMMs for recognition and matching. However, the study did not provide any testing results.

In [14], the authors center on the examination and identification of classical Arabic vocal phonemes, specifically vowels, through the utilization of HMM. They aim to tackle the issue of semantic changes that can occur due to variations in vowel durations (short or long) in Arabic. To investigate this, they examine three chapters (Alfateha, Albaqarah, and Alshuraa) from the Holy Quran. Their findings demonstrate an impressive overall accuracy rate of 87.60% without the utilization of a specified language model.

In [58], researchers developed a speech recognition system that utilizes MFCC for feature extraction and HMM for classification. The system focuses on recognizing and identifying the rules of Iqlab on Qira'at of Warsh. It uses a database of expert teachers' rules to compare and report any mismatches in the iqlab rules for specific verses. The system achieved a 70% accuracy in correctly spelling words with the correct rules from the database, a 50% accuracy for words with incorrect rules from the database, and a 40% accuracy for new words not included in the training database.

In [57], the researchers presented an interactive Tajweed system that assists in verifying the appropriate Imaalah Checking rule for Warsh recitation. This system utilizes an auto-

TABLE II. Summary of End-to-End based Speech Recognition in Written and Spoken Languages

Ref	Lang.	Main idea	FE	E2E DL	LM	AM
[29]	Arabic &Non-Arabic	A new strategy for multilingual ASR speech recognition. Implementation of three strategies to identify code-switching speech.	MFCCs and FB	Transformer based E2E Architecture	n-gram	TDNN & Transformer based E2E architecture
[31]	Arabic	A thorough examination to compare the E2E transformer ASR, the modular HMM-DNN ASR, and HSR.	MFCCs and Mel-spectrogram.	E2E transformer with hybrid (CTC+Attention)	LSTM and transformer-based language model (TLM)	combining a TDNN with LSTM layers.
[16]	Arabic	Build a robust diacritised Arabic ASR	MFCCs and the log Mel-Scale Filter Bank energies.	using joint CTC attention and using CNN-LSTM with attention	built by CMUCLMTK tool based on the 3-g and trained on RNN-LM.	KALDI, ESPnet, and Espresso
[9]	Arabic	E2E model for Arabic speech-to text transcription system using the lexicon free RNNs and CTC objective function based on Stanford CTC source code.	FB	BDRNNs	n-gram	TDNN-LSTM & BDRNNs
[25]	Non-Arabic	E2E training of AM using the LF-MMI objective function in the context of HMMs	MFCC	TDNN-LSTM	n-gram & RNN	E2E LF-MMI
[18]	Non-Arabic	Improving the performance of LAS which is a novel technique in ASR research.	FB	LAS ASR	5-gram	LAS ASR
[55]	Non-Arabic	Proposed purely E2E speech recognition open source framework called ESPnet toolkit.	MFCC & FB	E2E SR framework	RNN LM	CTC-objective fucnction

TABLE III. End-to-End based Speech Recognition Performance Compared to baseline Traditional Techniques

Ref.	Lang.	Dataset	Baseline Traditional Tech.		End-to-End Tech.	
			Model	WER/CER	Architecture	WER/CER
[29]	Arabic& non-Arabic	Arabic MGB2 &English TEDLIUM-3 & ES-CWA Corpus	Hybrid ASR	9.8%	E2E-Transformer	**8.29**%
[31]	Arabic	MGB2, and (Hidden Test (HT))	HMM-DNN	HT:15.9% MGB2: 15.8%	E2E-T (CTC + Attention)	HT:**12.6**% MGB2: **12.5**%
[16]	Arabic	Standard Arabic Single Speaker Corpus (SASSC)	Kaldi toolkit using DNN, MPE, and SGMM	33.72%	CNN-LSTM with attention using Espresso toolkit	**28.48**%
[9]	Arabic	8 hours Aljazeera corpus 1200 hours of TV Aljazeera corpus	TDNN-LSTM-BLSTM	14.7%	BDRNN with CTC objective function	**12.03**%
[55]	non-Arabic	Corpus of Spontaneous Japanese (CSJ)	HMM/DNN (Kaldi nnet1)	eval1:9.0% eval2:7.2% eval3:9.6%	ESPnet (i.e., VGG2-BLSTM, char-RNNLM, and joint decoding)	eval1:**8.7**% eval2:**6.2**% eval3:**6.9**%
[25]	non-Arabic	Switchboard And WSJ	Regular LF-MMI	Switchboard: **9.1**% WSJ: **2.8**%	E2E-LF-MMI	Switchboard: 9.6% WSJ:3.0%
[18]	non-Arabic	12,500 hour training set consisting of 15 million English utterances	hybrid HMM-LSTM	6.7% dictation task 5%	LAS end-to-end model	**5.6**% dictation task **4.1**%

matic speech recognition system with MFCC as the feature extraction technique and HMM as the classification method. The researchers conducted experiments using fifteen speech samples. The results showed that the system achieved a 60% accuracy rate for identifying the Imaalah rule based on the Warsh narration in the training data.

Researchers developed a system in [10] that identifies the Ahkam Al-Tajweed in a specific audio recording of Quranic recitation. The study focused on eight rules: “EdgamMeem” (one rule), “EkhfaaMeem” (one rule), “Ahkam Lam” in ‘Allah’ Term (two rules), and “Edgam Noon” (four rules). The classification problem involved 16 classes, covering the entire Holy Quran for verses that contained the eight rules. The system utilized various feature extraction techniques, including traditional methods like MFCC and LPC as well as newer methods like CDBN. Classifiers such as SVM and RF were employed, with the best accuracy of 96.4% achieved using SVM for classification and features extracted through MFCC, WPD, HMM-SPL, and CDBN.

In [59], authors developed a speech recognition system that can accurately differentiate between different types of Madd (elongated tone) and Qira’at (method of recitation) related to Madd. The system utilized MFCC as a feature extraction technique and HMM as a classification method. The focus of the study was on two specific types of Madd: greater connective prolongation and exchange prolongation rules for Hafss and Warsh. We collected a total of sixty data samples for analysis. The results showed that the accuracy of identifying the exchange prolongation rule was 60% for Warsh and 50% for Hafss. Additionally, the accuracy for identifying the greater connective prolongation rule was 40% for Warsh and 70% for Hafss.

Researchers developed an automated self-learning system in [33] to support the traditional method of teaching and learning Quran. The system aimed to classify the characteristics of Quranic letters. The study collected audio data from 30 participants, including 19 males and 11 females. The participants recited each sukoon alphabet once without repetition. The system used the Sukoon alphabet from the Quran to provide a description of the Makhraj (point of articulation) and Sifaat (characteristics) of each letter. The study successfully identified and classified the characteristic features of the alphabet, specifically in terms of learning (Al-Inhiraf) and repetition (Al-Takrir). The results showed that using QDA with all 19 features achieved the highest accuracy, with 82.1% for leaning (Al-Inhiraf) and 95.8% for repetition (Al-Takrir) characteristics.

The researchers conducted the study with the aim of creating a comprehensive system that accurately recognizes and determines the correct pronunciation of different Tajweed rules in audio. To achieve this, the researchers employed 70 filter banks as a feature extraction technique and utilized SVM as the classification method. The study focused on four specific rules, namely Ekhfaa Meem, Edgham Meem, Takhfeef Lam, and Tarqeeq Lam. The study utilized a dataset of 80 records, comprising a total of 657 recordings, encompassing both correct and incorrect recitations for each rule. They tested the models in the system against 30% of the recorded data and achieved a validation accuracy of 99%. [38] developed a recognition model to identify the “Qira’ah” from the corresponding Holy Quran acoustic wave. The study utilized MFCC as the feature extraction technique and SVM as the classification method. With a dataset of 258 wave files for 10 “Qira’ah” and including various reciters, the SVM accuracy achieved approximately 96%, while the accuracy of the ANN was 62%.

In another study, [45] proposed a system that automates the process of checking Tajweed for children who are learning the Quran. The system used the MFCC algorithm to extract the input speech signal and the HMM algorithm to compare children’s recitation with the recitation stored in the database. However, this project focused solely on Surah Al-Fatihah and did not provide any testing results.

Table IV summarizes the previous studies of template-based speech recognition in the HQSR field.

B. Traditional Based Speech Recognition

This section summarizes the papers that follow traditional speech recognition as described in Section II-A.

The researchers aimed to develop a precise Arabic recognizer for educational purposes in the study conducted by [34]. They implemented an HSMM model with the primary objective of improving the durational behavior of the traditional HMM model. To achieve this, they utilized a corpus consisting of recordings from 10 reciters, totaling over 487 minutes of speech. They meticulously segmented the corpus at three levels: phoneme, allophone, and words, with precise time boundaries. They obtained the recordings by reciting the Holy Quran, covering all the essential Arabic sounds. As a result of their work, the recognition accuracy saw an improvement of approximately 1.5%.

The researchers constructed an acoustic model using the Carnegie Melon University (CMU) Sphinx trainer [21]. The CMU Sphinx trainer utilized recordings from 39 different reciters and 49 chapters (surah) to build a robust framework for continuous speech recognition. The acoustic model achieved an impressive WER of approximately 15%, showcasing its accuracy and effectiveness.

In their research, [50] utilized data from everyayah.com, a website that provides open-access Quran recitations by numerous professional reciters, including Sheikh [1]. They adopted a deep learning approach to train an acoustic model for Quranic speech recognition. The study focused on 13 different reciters and concluded that the hybrid HMM-BLSTM method outperformed the HMM-GMM method in terms of speech recognition accuracy. The baseline models (HMM-GMM) achieved an average WER of 18.39%. In contrast, the acoustic model using Hybrid HMM-BLSTM achieved significantly better results, with an average WER of 4.63% in the same testing scenario.

The researchers utilized the KALDI toolkit to create and assess a speaker-independent continuous speech recognizer specifically designed for Holy Quran recitations in a study [49]. The researchers successfully developed a large-vocabulary system capable of recognizing and analyzing Quranic recitations. They use 32 recitations for Chapter 20 (Sūrat Taha), according to Hafs from the A’asim narration. The most effective experimental configuration involves utilizing Time Delay Neural Networks (TDNN) with a sub-sampling

TABLE IV. SUMMARY OF HQSR TEMPLATE-BASED SPEECH RECOGNITION

Ref#	Main Idea	Dataset	FE	ML algo.	Pros	Cons
[45]	Automated tajweed Checking System for Children.	Surah Al-Fatihah. Ten respondents' recitation for testing purposes. One audio of correct recitation is used for comparison with the respondents' audio	MFCC	HMM	-	very small data set & no testing result
[46]	The objective of this research was to differentiate between reliable and unethical Qur'anic reciters.	Seven well-known Qur'anic reciters have been gathered into a dataset. On an audio file, each reciter recited the Quran's surahs for eighty minutes.	MFCC	CNN	The proposed system stages are well-organized and easily understandable.	-
[38]	A recognition model for the "Qira'ah" from the corresponding Holy Quran acoustic wave.	The corpus contains 258 wave files labeled based on the "Qira'ah" (they consider 10 "Qira'ah") .	MFCC	SVM	good accuracy (96.12 %)	-
[11]	A system for recognizing/correcting the different rules of Tajweed in an audio.	Almost 80 records for each rule name and type, a total of 657 recordings of 4 different rules.	FBs	SVM	Good process for data collection.	Consider only 4 rules
[33]	Features identification and classification of alphabet (ro) in Leaning (Al-Inhiraf) and Repetition (AlTakrir) characteristics.	30 reciters(19 males and 11 females).	PSD & MFCC	LDA & QDA	Used multiple features extraction and classification methods	Not determined the used dataset
[59]	recognize, identify, and highlight discrepancies between two specific types of Madd rules: the greater connective prolongation and the exchange prolongation rules. This system focuses on verses that contain both rules and aims to point out the mismatches and differences between the rules for Hafss and Warsh recitation styles.	Reciter's database selected from Internet (60 data samples).	MFCC	HMM	-	few data samples
[10]	A system that determines which tajweed rule is used in a specific audio recording of a Quranic recitation (8 tajweed rules).	3,071 audio files collected from ten different expert reciters (5 males and 5 females). Each file contains a recording of one of the 8 rules considered (in either the correct or the incorrect usage).	Traditional (MFCC, LPC, WPD, HMM-SPL) & Non-traditional (CBDN)	KNN, SVM, ANN, RF, multiclass classifier, bagging.	Use of multiple feature extraction algorithms.	-
[57]	A system for distinguishing, recognizing, and correcting the pronunciation of tajweed rules for Warsh narration type).	15 speech simples. 5 verses recited by 3 Warsh .	MFCC	HMM	-	Few data samples
[58]	A system for recognizing, identifying and pointing out the mismatch of the iqlab rules for the verses containing the rules.	6 verses recited by 4 reciters with Qira'at of Warsh. Hence. The total is 24 of speech simples.	MFCC	HMM	-	Few data samples
[14]	differentiate between short and long vowels in Arabic. This distinction is crucial as it plays a significant role in altering the meaning of words.	MFCCs, deltas coefficients, deltas-deltas coefficients and the cepstral pseudoenergy	HMM	have a good accuracy	-	
[37]	A model to identify errors in the Quranic audio files and subsequently distinguish incorrect recitation from the correct recitation.	Ten of expert Qari each of them recite surat Al-Nnass ten times correct and ten times with mistakes.	MFCC	HMM and DTW	-	Covers only one sura
[6]	A system implemented using ASR technique.	Alnass, 20 utterances recited by only a one speaker with and without errors in recitation.	MFCC	ANN	-	Covers only one sura

technique. This setup achieved a WER ranging from 0.27% to 6.31% and a sentence error rate (SER) ranging from 0.4% to 17.39%.

The researchers of [3] used MFCC for the purpose of feature extraction. The researchers adjusted these features using the minimal phone error (MPE) as a discriminative model. The researchers utilized the deep neural network (DNN) model to construct the acoustic model. Here, they introduce an n-gram LM. The dataset utilized for training and assessing the proposed model comprises 10 hours of.wav recitations conducted by 60 reciters. The experimental results demonstrated that the proposed DNN model attained a remarkably low CER of 4.09% and a WER of 8.46%.

Table V summarizes the previous studies of traditional-based speech recognition in the HQSR field.

C. Other HQSR Studies

This section summarizes papers that follow other HQSR techniques, such as using the Google Speech API, Genetic Algorithm (GA), and MFCC.

The authors in [56] used MFCC to detect and recognize sounds for simple IDHAR tajweed without providing any testing results for this study.

Researchers proposed a solution in [22] to facilitate the memorization and learning of the Holy Quran. They employed the Fisher-Yates Shuffle algorithm to randomize the letters of the Quran, aiding in the memorization of verses. In addition, they employed the Jaro-Winkler algorithm for text matching and utilized the Google Speech API for speech recognition. The study focused on data from Juz 30. The achieved accuracy was approximately 91%, with an average matching time of 1.9 ms. However, the study revealed that it was still not possible to distinguish certain Arabic letters with similar pronunciations in Quranic verses in detail.

In [8], the authors produce a new speech segmentation algorithm for the Arabic language. Developing robust algorithms to accurately segment speech signals into fundamental units, rather than just frames, is a crucial preprocessing step in speech recognition systems. They focus on the precise segmentation of Quran recitation using multiple features (entropy, crossings, zero, and energy) and a GA-based optimization scheme. The results of the testing demonstrate a significant enhancement in segmentation performance, with an approximate 20% improvement compared to conventional segmentation techniques based on a single feature.

In [5], the authors implement an Android-based application called TeBook and provide a method for the assessment of the Holy Quran's recitation without the involvement of a third party by taking advantage of the use of speech recognition and an online Holy Quran search engine. There is no testing result present for this study. The limitation of this application is its reliance on multiple online services, which renders it unusable if the services are down.

The authors in [35] suggested a brand-new method called Samee'a to make it easier to memorize any form of literature, including speeches, poetry, and the entire Holy Qur'an. Samee's system utilizes the Jaro Winkler Distance technique to calculate the degree of similarity between the original and transformed texts, and employs the Google Cloud Speech Recognition API to translate Arabic speech to text. Seventy gathered files, ranging in length from twelve to four hundred words, together with a few chapters from the Holy Qur'an, were used to test the system. For the 70 files, the average similarity was 83.33%, while for the chosen chapters of the Holy Qur'an, it was 69%. Preprocessing operations on the text files and the Holy Qur'an improved these results to 91.33% and 95.66%, respectively.

In [48], the authors focus on the digital transformation of Quranic voice signals and the identification of Tajweed-based recitation faults in Harakaat as the primary research objective. They wanted to look into how to process speech using Quranic Recitation Speech Signals (QRSS) in the best digital format possible, using Al-Quran syllables and a design for feature extraction. The objective was to identify similarities or differences in recitation (based on Al-Quran syllables) between experts and students. We employ the DTW approach as a Short Time Frequency Transform (STFT) to quantify the Harakaat of QRSS syllable features. The research presents a method that utilizes human-guidance threshold classification to assess Harakaat, focusing on the syllables of the Qur'an. The categorization performance achieved for Harakaat exceeds 80% in both the training and testing phases.

Table VI summarizes the previously discussed studies of the other techniques used in the HQSR field.

D. HQSR Taxonomy

This section classifies the previously mentioned works of HQSR based on feature extraction methods and classification techniques. It is worth noting that most work done uses MFCCs as feature extraction techniques and the HMM as a classifier (e.g., [37], [14], [58], [57], [59], [45]). Fig. 4 illustrates the techniques of feature extraction and classification used in current HQSR research.

Researchers improved an Arabic recognizer by incorporating a HSMM instead of the traditional HMM [34]. Another approach, mentioned in [6], replaced HMM with ANN. Similarly, Nahar et al. [38] opted for SVM instead of HMM, and in [46] they use CNNs. Furthermore, researchers also explored various feature extraction techniques. For instance, [11] utilized FB for feature extraction and SVM for classification. [33] also used different methods, such as Formant Analysis, Power Spectral Density (PSD), and MFCC, along with LDA and QDA for sorting.

In [10], two categories of feature extraction techniques were employed: traditional and non-traditional. The traditional approach involved the utilization of MFCC, LPC, multi-signal WPD, and HMM-SPL. As for the non-traditional type, they use CDBN. They use K-Nearest Neighbors (KNN), SVM, ANN, Random Forest (RF), multiclass classifiers, and bagging for classification. [50] used MFCC for feature extraction, BLSTM as one of the deep learning topologies, and combined it with HMM as a hybrid system. The entire speech recognition system was built using the Kaldi toolkit [43], starting with feature extraction, acoustic modeling, and model testing. [49] used the deep learning approach in the KALDI toolkit to design, develop, and evaluate an ASR engine for the Holy

TABLE V. SUMMARY OF HQSR TRADITIONAL-BASED SPEECH RECOGNITION

Ref#	Main Idea	Dataset	FE	AM	LM	Pros	Cons
[49]	create a speech recognition engine that is independent of speaker and capable of handling continuous speech. Additionally, a written corpus that accurately represents the script of The Holy Quran is developed. To aid in the recognition process, a phonetic dictionary for The Holy Quran recitations is also be constructed.	32 recitations for Sūrat Taha according to Hafs from A'asim narration.	MFCC	KALDI toolkit to train the acoustic model (traditional and DNN approaches)	n-gram	Best research of current HQSR researches	Used only one sura
[50]	The acoustic model for Quran speech recognition was trained using a deep learning approach. In addition, the model was built to analyze the effect of Quran recitation styles (Maqam) on speech recognition.	The dataset is from everyayah.com [1].	MFCC	Hybrid HMM-BLSTM	3-grams	First work that used BLSTM in HQSR	no preprocessing method to eliminate noise and echo and not specified the used dataset
[21]	Used CMU Sphinx which is a robust framework for speaker-independent continuous speech recognition to train accurate acoustic models.	49 chapters were used: From chapter 067 to chapter 114 in addition to the chapter 001 and the supplication that is recited before the Holy Quran (Isti'adah).	CMU Sphinx framework			Get WER around 15% of trained acoustic model.	-
[34]	Presented the results of an enhanced Arabic recognizer by implementing an HSMM model instead of the standard one utilized in the baseline recognizer.	Arabic database utilized consists of 5935 waveform files for 10 reciters.	MFCC	HSMM	flat LM	Get enhancement by around 1.5% in the accuracy	-
[3]	Suggested the traditional method to recognizing Qur'an verses using a dataset of Qur'an verses.	A total duration of 10 hours of MP3 recordings containing recitations of Qur'an verses by 60 reciters.	MFCC	DNN	n-gram	Well-structured and clear article.	-

TABLE VI. SUMMARY OF OTHER HQSR STUDIES

Ref#	Main Idea	Dataset	Used Algorithms	Pros	Cons
[5]	Allow learners to learn how to memorize without the constraints of being in a fixed place and outside the classroom.	Everyayah.com Recitation Audio, Surah.my Translation	Alfanous JOS2 API, Android Speech Recognition	-	Relied heavily on online services and not specified the used dataset
[8]	A novel speech segmentation algorithm for Arabic language with a focus on the accurate segmentation of Quran recitation. Starting with a set of initial segmentations, three basic speech features: zero crossings, entropy, and energy are used.	They used the comprehensive KACST dataset with manually labelled Quran syllable structures.	feature fusion and Genetic Algoritms.	First segmentation on Quran recitation.	-
[22]	A solution to memorize and learn the Holy Quran easily.	juz 30	Fisher-Yates Shuffle Jaro-Winkler	-	Relies on Google Speech API which not trained on Quran verses
[56]	Emphasize idhar, which had a distinct and unambiguous pronunciation. The chosen hijaiyah letters comprised six possibilities, with only nun sukun and tanwin, making it effortless to identify them.	-	MFCC, FFT	-	Not specifying dataset, no testing result, and the poor organization of the content.
[35]	This article introduces a novel system called Samee'a, which aims to enhance the process of memorizing various types of texts, including poems, speeches, and the Holy Qur'an.	The system completed testing utilizing a dataset of 70 files, with word counts ranging from 12 to 400, including selected chapters from the Holy Qur'an.	Google Cloud Speech Recognition API and Jaro Winkler Distance algorithm	A comprehensive and informative paper	-
[48]	This study focuses on the digital transformation of Quranic voice signals and the identification of Tajweed-based recitation faults of Harakaat as its main research objective.	-	Dynamic Time Warping (DTW)	-	Not specifying dataset

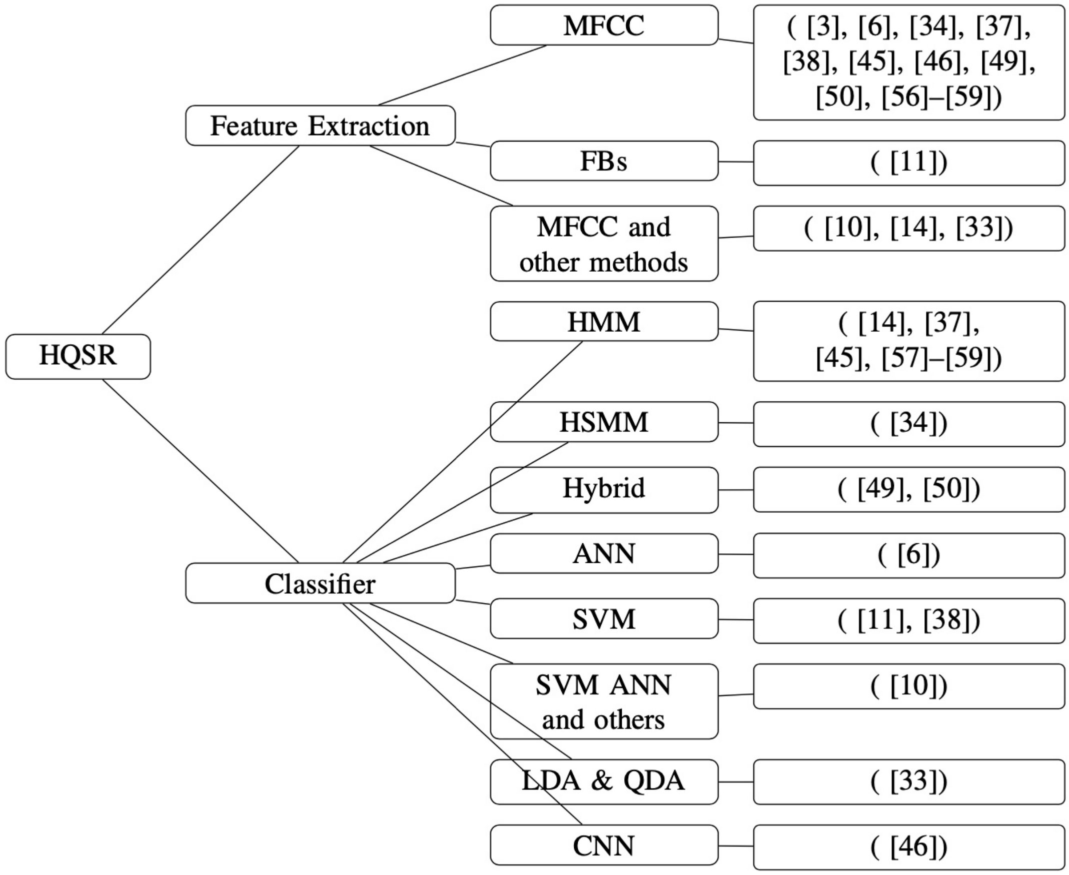


Fig. 4. HQSR Taxonomy of Used Technique

Quran recitations. The best experimental setup was achieved using TDNN with sub-sampling technique.

In [21], the CMU Sphinx trainer [53] was employed to train the acoustic model specifically for the Holy Quran. In a similar vein, a study by [22] utilized the Jaro-Winkler algorithm for text matching and relied on the Google Speech API to establish a framework for speech recognition.

The solution of [5] uses Android speech recognition and depends heavily on third-party online services. In [8], they developed a robust hybrid speech segmentation system based on multiple features (entropy, zero crossings, and energy) and a GA-based optimization scheme to obtain accurate segment units specially adapted for Quran recitation.

IV. Discussion and Future Research Directions

Recent research in the field of HQSR has suggested numerous works. We provide in Table VII a comparative analysis of current research works based on the dataset characteristics and the suggested methodology:

- Dataset characteristics:
  1) #verses: This refers to the total number of verses used in the study: L (number of verses between 1-100), M (number of verses between 101-200), H (number of verses greater than 200), and N (number of verses not determined in the paper).
  2) #sura: This refers to the total number of suras used in the study.
  3) #reciters: This shows the number of reciters who participate in this study.
- Proposed methodology:
  1) DL-based: This criteria shows if the study used deep learning in any stage of the study.
  2) LM: This shows if the study used a language model in their solution or not.
  3) AM: This shows if the study used an acoustic model in their solution or not.
  4) Reciter Independent: This shows if the output model of this study is reciter independent or not.
  5) Speaker Adaptation: This shows if this study used any techniques of speaker adaptation or not.

As we can see in Table VII, there is a significant gap in the current work of HQSR. First, most of the works in HQSR follow template-based speech recognition, which is an old style of speech recognition. This style extracts features of raw audio and feeds these features into a classifier to classify and match with stored templates without using acoustic, lexical (pronunciation), or language models [58]. Examples of these works are ([45], [38], [11], [33], [59], [57], [58], [14], [6], and [37]) as shown in Table VII. Few works suggest the use of deep learning. However, they still rely on an old-style design. For instance, [10] used a deep learning architecture with the old style (i.e., it didn't use acoustic, lexical, and language models but deep learning in the feature extraction phase only). In addition, the authors in [5], [22] employed the Google Speech API for their solution. However, this approach had limitations as the API was unable to accurately differentiate between the seven letters that share similar pronunciations in the verses of the Quran.

Second, only a few studies follow traditional speech recognition, as explained in Fig. 2, either with a deep learning architecture like [49], [50], [3] or without, such as [34], [21]. It is worth noting that more work investigating deep learning architecture should be conducted to improve the accuracy of Arabic speech recognition in general and the Holy Quran in particular.

Third, we can observe in Table VII that the used data set is too small for most of the work (number of verses, suras, and reciters). while an extensive dataset helps produce a robust and generalized speech recognition system.

Finally, no research on HQSR used end-to-end deep learning architecture, while this architecture shows outstanding results with Arabic and non-Arabic languages, as previously discussed in Section II-B (see Table III, which presents a comparison between some end-to-end based speech recognition architecture and traditional techniques in Arabic and non-Arabic languages).

To sum up, many challenges still need to be considered in future work. Indeed, in recognition of speech, recognizing individual words is easy, but the challenge is recognizing continuous speech [7]. Multiple conditions, including speaker dependency, vocabulary size, and noisy environments, can affect the performance of speech recognition systems. Recognition performance increases with limited vocabulary and speaker-dependent conditions while using broad vocabulary and speaker-independent scenarios; performance can decrease significantly [7]. Moreover, Arabic is a morphologically complex language that contains a high degree of affixation and derivation, resulting in a massive increase in word forms [31]. Furthermore, speech recognition of the Holy Quran has additional difficulties compared with written and spoken languages for the following reasons:

- Lack of a comprehensive dataset that contains recitations of women, children, and native and non-native Arabic speakers with both the correct and incorrect recitation of the Holy Quran.
- Mistakes are not acceptable when reading the Quran because an error in reciting only one letter may change the meaning.
- The diversity of narrations in reading the Qur'an makes it difficult for the model to recognize different narrations.
- The diversity of *Magam* in Quran Recitation, such as (*bayat*, *Ajam*, *Nahawand*, *Hijaz*, *Rost*, *Sika*, etc.), adds more difficulty for the model when recognizing the recitation.
- The length of prolongation (*Madd*) varies when reciting the Quran. In Hafs An Asim narration, reciters can recite some types of the madd with 2, 4, or 5 *Harakat*.
- Recitation of the Holy Quran must follow the rules of "tajweed" and correctly pronounce Makhraj (point of articulations) and the Sifaat (characteristics) of each alphabet.

TABLE VII. COMPARING HQSR SOLUTIONS

Ref#	Dataset			Methodology				
	#verses	#sura	#reciters	DL-based	LM	AM	Reciter Independent	Speaker Adaptation
[45]	L	1	1					
[49]	M	1	32	✓	✓	✓	✓	✓
[38]	H							
[11]	H							
[33]	N		30					
[5]	N							
[50]	N		13	✓	✓	✓		✓
[22]	H							
[59]	L							
[10]	H		10	✓				
[57]	L							
[58]	L							
[56]	N							
[21]	H	49	39		✓	✓	✓	
[34]	H		10		✓	✓		
[14]	H	3	4					
[37]	L		10					
[6]	L	1	1					
[46]	H		7	✓				
[3]	L		60	✓	✓	✓		

Note: *L (number of verses between 1-100), M (number of verse between 101-200), H (number of verses greater than 200), and N (Number of verses not determined in the paper).*

V. CONCLUSION

This paper surveys Holy Quran Speech Recognition (HQSR) works. It summarizes some studies of speech recognition in written and spoken languages and the most recent work in the HQSR field. It provides a general taxonomy of speech recognition and a specific one dedicated to HQSR studies that illustrates the techniques of feature extraction and classification used in current HQSR research. We compared the current solutions and clarified the limitations of the current studies. The main challenges of the HQSR field are the lack of a comprehensive dataset, minimizing mistakes that are not acceptable when reading the Quran, diversity of narrations, diversity of Magam in Quran recitation, and diversity of prolongation (Madd) length when reciting the Quran. The field of HQSR needs a lot of work to improve the current speech recognition models of the Holy Quran by using better techniques that already show good results with written and spoken languages but haven't been used with HQSR yet.

REFERENCES

[1] Every ayah, http://www.everyayah.com/, 2022.

[2] Muslim population by country 2021, https://worldpopulationreview.com/country-rankings/muslim-population-by-country, 2021.

[3] Alsayadi Hamzah A and Hadwan Mohammed. Automatic speech recognition for qur'an verses using traditional technique. *Journal of Artificial Intelligence and Metaheuristics (JAIM)*, 2022.

[4] Abdelaziz A Abdelhamid, Hamzah A Alsayadi, Islam Hegazy, and Zaki T Fayed. End-to-end arabic speech recognition: A review. 2020.

[5] Mohd Hafiz Bin Abdullah, Zalilah Abd Aziz, Rose Hafsah Abd Rauf, Noratikah Shamsudin, and Rosmah Abd Latiff. Tebook a mobile holy quran memorization tool. In *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pages 1–6. IEEE, 2019.

[6] Bushra Abro, Asma Batool Naqvi, and Ayyaz Hussain. Qur'an recognition for the purpose of memorisation using speech recognition technique. In *2012 15th International Multitopic Conference (INMIC)*, pages 30–34. IEEE, 2012.

[7] Ahmed Hamdi Abo Absa. *Self-Learning Techniques for Arabic Speech Segmentation and Recognition*. Thesis, 2018.

[8] Ahmed Hamdi Abo Absa, Mohamed Deriche, Moustafa Elshafei-Ahmed, Yahya Mohamed Elhadj, and Biing-Hwang Juang. A hybrid unsupervised segmentation algorithm for arabic speech using feature fusion and a genetic algorithm (july 2018). *IEEE Access*, 6:43157–43169, 2018.

[9] Abdelrahman Ahmed, Yasser Hifny, Khaled Shaalan, and Sergio Toral. End-to-end lexicon free arabic speech recognition using recurrent neural networks. *Computational Linguistics, Speech And Image Processing For Arabic Language*, pages 231–248, 2019.

[10] Mahmoud Al-Ayyoub, Nour Alhuda Damer, and Ismail Hmeidi. Using deep learning for automatically determining correct application of basic quranic recitation rules. *Int. Arab J. Inf. Technol.*, 15(3A):620–625, 2018.

[11] Ali M Alagrami and Maged M Eljazzar. Smartajweed automatic recognition of arabic quranic recitation rules. *arXiv preprint arXiv:2101.04200*, 2020.

[12] Abdulmalik A Alasadi, TH Aldhayni, Ratnadeep R Deshmukh, Ahmed H Alahmadi, and Ali Saleh Alshebami. Efficient feature extraction algorithms to develop an arabic speech recognition system. *Engineering, Technology & Applied Science Research*, 10(2):5547–5553, 2020.

[13] Hanan Aldarmaki, Asad Ullah, and Nazar Zaki. Unsupervised automatic speech recognition: A review. *arXiv preprint arXiv:2106.04897*, 2021.

[14] Yousef A Alotaibi, Mohammed Sidi Yakoub, Ali Meftah, and Sid-Ahmed Selouani. Duration modeling in automatic recited speech recognition. In *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, pages 323–326. IEEE, 2016.

[15] Fatimah Alqadheeb, Amna Asif, and Hafiz Farooq Ahmad. Correct pronunciation detection for classical arabic phonemes using deep learning. In *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*, pages 1–6. IEEE, 2021.

[16] Hamzah A Alsayadi, Abdelaziz A Abdelhamid, Islam Hegazy, and Zaki T Fayed. Arabic speech recognition using end-to-end deep learning. *IET Signal Processing*, 2021.

[17] Dario Amodei, Sundaram Ananthanarayanan, Rishita Anubhai, Jingliang Bai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Qiang Cheng, and Guoliang Chen. Deep speech 2: End-to-end speech recognition in english and mandarin. In *International conference on machine learning*, pages 173–182. PMLR, 2016.

[18] Chung-Cheng Chiu, Tara N Sainath, Yonghui Wu, Rohit Prabhavalkar, Patrick Nguyen, Zhifeng Chen, Anjuli Kannan, Ron J Weiss, Kanishka Rao, and Ekaterina Gonina. State-of-the-art speech recognition with sequence-to-sequence models. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4774–4778. IEEE, 2018.

[19] Ronan Collobert, Awni Hannun, and Gabriel Synnaeve. A fully differentiable beam search decoder. In *International Conference on Machine Learning*, pages 1341–1350. PMLR, 2019.

[20] Ronan Collobert, Christian Puhrsch, and Gabriel Synnaeve. Wav2letter: an end-to-end convnet-based speech recognition system. *arXiv preprint arXiv:1609.03193*, 2016.

[21] Mohamed Yassine El Amrani, MM Hafizur Rahman, Mohamed Ridza Wahiddin, and Asadullah Shah. Towards an accurate speaker-independent holy quran acoustic model. In *2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pages 1–4. IEEE, 2017.

[22] YA Gerhana, AR Atmadja, DS Maylawati, A Rahman, K Nufus, H Qodim, and MA Ramdhani. Computer speech recognition to text for recite holy quran. In *IOP Conference Series: Materials Science and Engineering*, volume 434, page 012044. IOP Publishing, 2018.

[23] Alex Graves, Santiago Fern?ndez, Faustino Gomez, and Jürgen Schmidhuber. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning*, pages 369–376, 2006.

[24] Imane Guellil, Houda Saâdane, Faical Azouaou, Billel Gueni, and Damien Nouvel. Arabic natural language processing: An overview. *Journal of King Saud University-Computer and Information Sciences*, 33(5):497–507, 2021.

[25] Hossein Hadian, Hossein Sameti, Daniel Povey, and Sanjeev Khudanpur. End-to-end speech recognition using lattice-free mmi. In *Interspeech*, pages 12–16, 2018.

[26] Awni Hannun, Ann Lee, Qiantong Xu, and Ronan Collobert. Sequence-to-sequence speech recognition with time-depth separable convolutions. *arXiv preprint arXiv:1904.02619*, 2019.

[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[28] Xiaodong He and Li Deng. Discriminative learning for speech recognition: theory and practice. *Synthesis Lectures on Speech and Audio Processing*, 4(1):1–112, 2008.

[29] Ahmed Ali Hifny, Shammur Absar Chowdhury, Amir Hussein, and Yasser. Arabic code-switching speech recognition using monolingual data. *Proc. Interspeech 2021*, pages 3475–3479, 2021.

[30] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, and Tara N Sainath. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6):82–97, 2012.

[31] Amir Hussein, Shinji Watanabe, and Ahmed Ali. Arabic speech recognition by end-to-end, modular systems and human. *arXiv preprint arXiv:2101.08454*, 2021.

[32] Biing Hwang Juang and Laurence R Rabiner. Hidden markov models for speech recognition. *Technometrics*, 33(3):251–272, 1991.

[33] Safiah Khairuddin, Salmiah Ahmad, Abdul Halim Embong, Nik Nur Wahidah Nik Hashim, and Surul Shahbuddin Hassan. Features identification and classification of alphabet (ro) in leaning (al-inhiraf) and repetition (al-takrir) characteristics. In *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pages 295–299. IEEE, 2019.

[34] Mohamed OM Khelifa, Mostafa Belkasmi, Yousfi Abdellah, and Yahya OM ElHadj. An accurate hsmm-based system for arabic phonemes recognition. In *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)*, pages 211–216. IEEE, 2017.

[35] Souad Larabi-Marie-Sainte, Betool S. Alnamlah, Norah F. Alkassim, and Sara Y. Alshathry. A new framework for arabic recitation using speech recognition and the jaro winkler algorithm. *Kuwait Journal of Science*, 49, 2022.

[36] Lina Marlina, Cipto Wardoyo, WS Mada Sanjaya, Dyah Anggraeni, Sinta Fatmala Dewi, Akhmad Roziqin, and Sri Maryanti. Makhraj recognition of hijaiyah letter for children based on mel-frequency cepstrum coefficients (mfcc) and support vector machines (svm) method. In *2018 International Conference on Information and Communications Technology (ICOIACT)*, pages 935–940. IEEE, 2018.

[37] Ammar Mohammed, Mohd Shahrizal Sunar, and Md Sah Hj Salam. Quranic verses verification using speech recognition techniques. *Jurnal Teknologi*, 73(2), 2015.

[38] Khalid MO Nahar, M Ra'ed, A Moy'awiah, and M Malek. An efficient holy quran recitation recognizer based on svm learning model. *Jordanian Journal of Computers and Information Technology (JJCIT)*, 6(04), 2020.

[39] Maryam Najafian, Wei-Ning Hsu, Ahmed Ali, and James Glass. Automatic speech recognition of arabic multi-genre broadcast media. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 353–359. IEEE, 2017.

[40] Maryam Najafian, Sameer Khurana, Suwon Shan, Ahmed Ali, and James Glass. Exploiting convolutional neural networks for phonotactic based dialect identification. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5174–5178. IEEE, 2018.

[41] Lv Ping. English speech recognition method based on hmm technology. In *2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pages 646–649. IEEE, 2021.

[42] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, and Petr Schwarz. The kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*. IEEE Signal Processing Society, 2011.

[43] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, and Petr Schwarz. The kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*. IEEE Signal Processing Society, 2011.

[44] Rohit Prabhavalkar, Kanishka Rao, Tara N Sainath, Bo Li, Leif Johnson, and Navdeep Jaitly. A comparison of sequence-to-sequence models for speech recognition. In *Interspeech*, pages 939–943, 2017.

[45] Munirah Ab Rahman, Izatul Anis Azwa Kassim, Tasiransurini Ab Rahman, and Siti Zarina Mohd Muji. Development of automated tajweed checking system for children in learning quran. *Evolution in Electrical and Electronic Engineering*, 2(1), 2021.

[46] Ghassan Samara, Essam Al-Daoud, Nael Swerki, and Dalia Alzu'bi. The recognition of holy qur'an reciters using the mfccs' technique and deep learning. *Advances in Multimedia*, 2023, 2023.

[47] Benjamin Elisha Sawe. Arabic speaking countries, Jul 2018.

[48] Noraimi Shafie, Azizul Azizan, Mohamad Zulkefli Adam, Hafiza Abas, Yusnaidi Md Yusof, and Nor Azurati Ahmad. Dynamic time warping features extraction design for quranic syllable-based harakaat assessment. *International Journal of Advanced Computer Science and Applications*, 13, 2022.

[49] Imad K Tantawi, Mohammad AM Abushariah, and Bassam H Hammo. A deep learning approach for automatic speech recognition of the holy qur'an recitations. *International Journal of Speech Technology*, pages 1–16, 2021.

[50] Faza Thirafi and Dessi Puji Lestari. Hybrid hmm-blstm-based acoustic modeling for automatic speech recognition on quran recitation. In *2018 International Conference on Asian Language Processing (IALP)*, pages 203–208. IEEE, 2018.

[51] Pahini A Trivedi. Introduction to various algorithms of speech recognition: Hidden markov model, dynamic time warping and artificial neural networks. *International Journal of Engineering Development and Research*, 2(4):3590–3596, 2014.

[52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, ?ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[53] Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, and Joe Woelfel. Sphinx-4: A flexible open source framework for speech recognition. 2004.

[54] Song Wang and Guanyu Li. Overview of end-to-end speech recognition. In *Journal of Physics: Conference Series*, volume 1187, page 052068. IOP Publishing, 2019.

[55] Shinji Watanabe, Takaaki Hori, Shigeki Karita, Tomoki Hayashi, Jiro Nishitoba, Yuya Unno, Nelson Enrique Yalta Soplin, Jahn Heymann, Matthew Wiesner, and Nanxin Chen. Espnet: End-to-end speech processing toolkit. *arXiv preprint arXiv:1804.00015*, 2018.

[56] Efy Yosrita and Abdul Haris. Identify the accuracy of the recitation of al-quran reading verses with the science of tajwid with mel-frequency ceptral coefficients method. In *2017 International Symposium on Electronics and Smart Devices (ISESD)*, pages 179–183. IEEE, 2017.

[57] Bilal Yousfi and Akram M Zeki. Holy qur'an speech recognition system imaalah checking rule for warsh recitation. In *2017 IEEE 13th international colloquium on signal processing & its applications (CSPA)*, pages 258–263. IEEE, 2017.

[58] Bilal Yousfi, Akram M Zeki, and Aminah Haji. Isolated iqlab checking rules based on speech recognition system. In *2017 8th International Conference on Information Technology (ICIT)*, pages 619–624. IEEE, 2017.

[59] Bilal Yousfi, Akram M Zeki, and Aminah Haji. Holy qur'an speech recognition system distinguishing the type of prolongation. *Sukkur IBA Journal of Computing and Mathematical Sciences*, 2(1):36–43, 2018.

[60] Naima Zerari, Samir Abdelhamid, Hassen Bouzgou, and Christian Raymond. Bidirectional deep architecture for arabic speech recognition. *Open Computer Science*, 9(1):92–102, 2019.

[61] Yaxin Zhang, Mike Alder, and Roberto Togneri. Using gaussian mixture modeling in speech recognition. In *Proceedings of ICASSP'94. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages I/613–I/616 vol. 1. IEEE, 1994.

Improving the Classification of Airplane Accidents Severity using Feature Selection, Extraction and Machine Learning Models

Rachid KAIDI[1], Mohammed AL ACHHAB[2], Mohamed LAZAAR[3], Hicham OMARA[4]
ENSA, Abdelmalek Essaadi University, Tetouan, Morocco[1,2]
ENSIAS, Mohammed V University, in Rabat, Morocco[3]
FS, Abdelmalek Essaadi University, Tetouan, Morocco[4]

***Abstract*—Airplane mode of transportation is statistically the most secure means of travel. This is due to the fact that flights require several conditions and precautions because aviation accidents are most of the time fatal and have disastrous consequences. For this purpose, in this paper, the mean goal is to study the different levels of fatality of airplane accidents using machine learning models. The study rely on airplane accident severity dataset to implement three machine learning models: KNN, Decision Tree and Random Forest. This study began with implementing two features selection and extraction methods, PCA and RFE in order to reduce dataset dimensionality and complexity of models and reduce training time by implementing machine learning models on dataset and measuring their performance. Results show that KNN and Decision Tree demonstrates high levels of performances by achieving 100% of accuracy and f1-score metrics; while Random Forest achieves its best performances after application of PCA when it reaches an accuracy equal to 97.83% and f1-score equal to 97.82%.**

Keywords*—Airplane accident; severity; flights safety; machine learning; KNN; Random Forest (RF); Decision Tree (DT)

I. Introduction

It is well known that the plane is the safest mode of transportation. In 2022, there have been only two fatal plane crashes without counting small and helicopter crashes [1]. In 2007, The National Transportation Safety Board known as NTSB, claimed that 24 million hours of plane travels had occurred, only 6.84 of every 100.000 flights hours had a plane accident and 1.19 of every 100.000 plane flights are fatal crashes. The aviation industry contribute significantly to the national economic of each country if they have robust and strong aviation policies and technologies. Thousands of incomes in this field can be reached every year; for this purpose, this industry is well developed and controlled by international standards. Among the most important requirements in aviation industry is safety. That is why different measures of security are taken into consideration before and after any flight. These requirements include:

- A strict safety requirements based of international standards in order to establish a base to rate degree of safety in any flight.
- Collecting data in order to perform data analytics to find out any shortcoming and to perform safety improvements.
- Continuous extensive training for pilots to update their knowledge and skills.
- Safety Management System (SMS) should be implemented in any plane in order to have synchronous state of the plane.
- Auditing safety measures by investigating incidents, analysing performances indicators, etc.

The most important requirement for us in this study is the use of data in order to perform safety audit, prevent accidents and rectify any breaches of policies or technologies misconfigurations or malfunctions. As we can say, safety requirements are very hard and complex to implement because any small error or misconfigurations may lead to fatal consequences. According to [13] airplane accidents can be caused by so many factors including: Pilot error due to miscommunication, distraction, exhaustion, drainage, etc. mechanical error, bad weather conditions, sabotage and human errors. So many scientific contributions had tried to implement data0based approaches using machine learning (ML) models to deal with these safety issues specially in the context of our study which is the prediction of severity of airplane incidents. [4,6,7,12,13] propose ML-based models, deep learning are used and implemented on complex dataset that need deep models such as in [11,14]. Authors using machines learning (ML) models achieved promising results but there is always some data and implementations constraints including limited resources and information about fatal accidents because there are very rare to occur; that is why it is difficult to collect enough data to establish meaningful statistical analysis. The factors that control the operations of an airplane are numerous and complex; they include environmental factors, techniques, human resources factors. Data may be biased toward the condition and purposes so it can misrepresent some conditions and important factors. Mathematical representation of severity may be very difficult for modelisation using classical approaches such as text mining [2]. Hence the need for machine learning solutions. Emre Kuşkapan et al. [8] propose an approach for aviation accidents classification using data mining algorithms. They collect data worldwide from 2000 to 2019. They implement J48, Naïve Bayes and Sequential Minimal Optimization methods. J48 outperforms all methods based on Precision, Recall, F-measure and ROC Area. L. J. Raikar et al. [3] implement SVM, KNN, Adaboost, XGboost to analyse airplane crash. They include feature selection and scaling methods in order to reduce dimensionality of data by removing unnecessary characteristics. In this study, we will implement machine learning models on airplane accident severity dataset to predict severity of airplane accidents. Before that, we established a robust phase of feature selection and extraction in order to get the most relevant and important features. These features will be the focus of future work. We get interesting results. Some ML models reached 100% of accuracy using KNN and Decision Tree. The rest of this paper is organised as follows: In Section 2, we will discuss the related work, in Section 3, we will present the background of this study and the followed methodology. In Section 4, we will present our results, discuss and criticise them and finally a conclusion where we will mention the relevant results, limits of this study and its perspectives in future work.

II. Related Work

R. A. Burnett et al. [6] implement machine learning models in order to predict the injuries and fatalities in aviation accidents. They face so many problems in data processing. First of all, they needed to deal with redundant fields, missing values, lack of generalisation which means that there are lot of changes of conditions over years, so implementing machine learning models on old statistics may lead to misrepresent the results. Another raised problem is that they needed to deal with it is imbalanced data which is a very complex task in machine learning context. They used six Federal Aviation Administration Aviation Incidents and accidents records in range from 1975-2002. They Implement KNN, SVM and KNN models to predict the rate of aviation accidents. Results show that ANN gives a promising results and its obvious because ANN models can generalise and analyse internal relationships between features in order to extract pattern more better than regular machine learning models.

It is well know that in any information system, the human being is the most vulnerable asset. Human factor contribute to approximatively 75% of aircraft accidents and incidents [5]. In the paper by M. Bagarzan et al. [7], they conducted an interesting study in order to analyse the impact of the age, experience and gender of pilots on aviation accidents. They specify six categories of age: “less than 20”, “20-29”, “30-39”, “40-49”, “50-59” and “more than 60”. “Experience” in this study is based on the number of hours for each pilot involved in an accident. the records belong to the NTSB database. They implement chi-square and logistic regression models in their study in order to figure out if there is any relationships between pilot characteristics and how much these characteristics can contribute in causing aviation accident that lead to serious consequences. Results show that the gender had no great impact of the pilot error but females make fatal errors less than men. Pilots that are older than 60 years old can make pilot error. Experienced pilots can easily get involved in fatal accidents because due to their experiences they can fly in conditions that non experienced pilots cannot do but these experienced pilots are less likely to make pilot errors. Authors suggest that there are environment conditions that can affect the performances of a pilot and they suggest to maintain training for pilots and improve performances of technologies used in this mode of transportation.

In the study by N. Pande et al. [13], they conducted a study about the prediction of fatal aviation accidents. They used Random Forest, XGBoost, Neural Network, multiple Linear Regression, chi-square, linear regression, ensemble model that combine so many machine learning models and logistic regression. This study is based on a data that is collected since 1908. 4700 data points of it were used in this study. What is interesting about this study is that it is based on simple classical machine learning models like RF, XGBoost and complex machine learning models that refers to the combination of different machine learning models in order to predict the dependant feature. The evaluation metrics in this study include minimum error, maximum error, mean absolute error, linear correlation and standard deviation. Results show that Neural Network model outperform all of the implemented models reaching an accuracy equal to 90.6% which is the case for [13].

We can say that machine learning models are widely used in the field of severity and fatalities of flights accidents. The usefulness of machine learning models in this field are related to the complexity of data. In order words, same conditions of weather and other plane characteristics that we can gather using sensors, the same data can lead to different results. We can never be sure of these characteristics because even a bird strike [9] can cause fatal damage.

III. Theoretical Background

In order to implement our ML models, we will rely on two models for data processing: PCA for feature extraction and RFE for features selection. Then we will implement our ML models that are: RF, DT and KNN.

A. Feature Extraction: PCA

It is an abbreviation for Principal Component Analysis [19]. It is an unsupervised ML model that is used to reduce data dimensionality based on statistical measurements, and generates new components to contain the most significant feature data by capturing a large amount of variance[20]. PCA is used to transform a large set of data into a small one but keeping relevant information about data. The principal components of PCA are orthogonal. PCA is useful to reduce the noise in data, to compress data and helps in visualising data with high dimensions and to detect any relationships between features mainly correlation and other relationships in order to gather correlated features with each others. Given a dataset X with n observations and p variables, we can perform PCA by following these steps:

Center the data by subtracting the mean $\bar{x}$ from each variable:

$$\tilde{x}ij = xij - \bar{x}j \tag{1}$$

where $\tilde{x}ij$ is the centered value of variable j in observation i.

Calculate the sample covariance matrix S:

$$S = \frac{1}{n-1}\sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^T \tag{2}$$

where $\tilde{x}_i$ is the centered vector of observation i.

Compute the eigenvectors $v_1, v_2, ..., v_p$ and corresponding eigenvalues $\lambda_1, \lambda_2, ..., \lambda_p$ of S.

Select the k eigenvectors with the highest eigenvalues, where k is the number of dimensions in the reduced feature space.

Project the centered data onto the k selected eigenvectors to obtain the reduced feature space:

$$Y = XV_k \tag{3}$$

where V_k is a matrix containing the top k eigenvectors as columns.

The final output Y will have dimensions $n \times k$, where each row represents an observation and each column represents a principal component.

B. Feature Selection: RFE

Recursive Feature Elimination is a feature selection method used to get the most important features that contribute to improve ML models performances. RFE is a recursive method that removes in each iteration the worst features. We first provide RFE with all features, RFE run on ML models on the dataset for instance Random Forest. After training the model -ML model- RFE measures the contribution and importance of each feature. Less relevant features will be removed and re-run the model until we get the best number of features that contribute the most to the model. The process of Recursive Feature Elimination (RFE) involves assigning weights to different features in order to identify which ones contribute the most towards predicting the target variable. This is done by ranking the features based on their relative importance [21], which will help to decrease complexity of the model, minimising time of training and increasing model performances. The RFE algorithm can be mathematically represented as follows: Let X be the feature matrix and y be the target vector. Let n be the total number of features and k be the desired number of features.

1) Initialize $X_{\text{RFE}} = X$ and $k_{\text{RFE}} = n$.
2) Train a model on X_{RFE} and y to obtain coefficients or feature importance.
3) Calculate the importance of each feature
4) Remove the least important feature from X_{RFE} to obtain X'_{RFE} and decrement k_{RFE}.
5) we repeat the same steps until $k_{\text{RFE}} = k$.

C. Decision Tree

It is a supervised ML model used for classification and regression.DT are not only highly useful in various applications but also renowned for their interpretability and resilience [16] . DT split the data based on features with most importance. These features importance is measured using Gini index or Entropy. The final structure of a DT model is a tree, where nodes represent features and leafs represent the dependant feature. The process of a DT model is as follow: The dataset went through DT from input node to the leafs where each leaf refer to a value of the dependant feature. Unlike RF, DT may have bad performances in front of high dimensionality datasets or noisy datasets.

To construct a Decision Tree, the ID3[15] algorithm is utilized, which involves several steps. The initial stage involves computing the entropy or Gini impurity of the target class in order to assess the data's impurity.

$$Gini(S) = 1 - \sum_{i=1}^{k} p_i^2 \tag{4}$$

Where k represents the total number of classes, while p_i denotes the proportion of instances that belong to the i-th class.

Afterward, we calculate the Gini gain for each attribute in our dataset and select the attribute that provides the greatest value and create a node for that attribute. The formula used to calculate the Gini gain for each attribute is:

$$Gini\_Gain(S, A) = Gini(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} Gini(S_v) \tag{5}$$

Where A represents an attribute, while S refers to the dataset. S_v represents the subset of instances in S where attribute A has a value of v. Repeat steps 2 recursively for every subset of data generated by the split until all instances within a subset are categorized under the same class or there are no remaining attributes left to split the data.

D. Random Forest

Random forest [17] is a supervised machine learning model used for many purposes including classification and regression. It is based on building Decision trees on subset of the samples and features. RF contains n DT. The choice of n depends on the task. Each DT is trained on a random part of the data using random partitions that helps to decrease the model's complexity and to prevent it from overfitting. Each DT makes a prediction and then the majority vote will be considered for prediction. RF gives best results on large data of high dimensions and RF is very useful in case of noisy data.

E. KNN

Abbreviation of K-Nearest Neighbors. It is a non-parametric supervised machine learning model [18], mainly used for classification and regression tasks. The purpose of KNN is to find the K nearest data points of data in order to make a prediction. KNN measures the distance between data points using Euclidean distance:

$$d\left(X, Y\right) = \sqrt{\sum_{i=1}^{n} \left(X_i - Y_i\right)^2} \tag{6}$$

Or Manhattan distance:

$$d\left(X, Y\right) = \sum_{i=1}^{n} |X_i - Y_i| \tag{7}$$

These distances are used to group subsets of data in order to measure the dependant variable in the training phase. In case of large datasets, KNN may need additional computational resources. The choice of the k is not trivial, choosing a wrong k will lead to performances degradation.

F. Accuracy

Accuracy is a commonly used metric in machine learning to evaluate the performance of a model. It involves counting the number of true positive (TP) and true negative (TN) samples in a given dataset, and dividing it by the total number of samples including false positive (FP) and false negative (FN) samples. In other words, accuracy measures how many samples are correctly predicted out of all samples in the dataset.

$$Accuracy = \frac{TP + TN}{TP + TF + FP + FN)} \tag{8}$$

G. F1-Score

The F1-score is a metric used to measure the classification performance of machine learning models. It combines two different metrics, precision and recall, to provide a single score that reflects the overall performance of the model. A good F1-score requires good results for both precision and recall, or high results for one metric if the other metric has low results, in order to balance the results.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{9}$$

IV. Results and Discussion

Table I shows the results of KNN, DT and Random Forest using all features. Results show that KNN and DT reached 100% of performances for both metrics: accuracy and f1-score. RF reaches an accuracy equal to 95.48% and f1-score equal to 95.49%.

TABLE I. KNN, DT and RF ML Models Performances using All Features based on Accuracy and F1-score

	Accuracy	F1-Score
Random Forest	95.48%	95.49%
Decision Tree	100%	100%
KNN	100%	100%

In order to reduce the dimensionality of the dataset, we used PCA with Principal Components equal to 8. Results in Table II show that results remain the same for KNN and DT, which means that PCA preserved relevant inertia of original dataset while reducing complexity of the models and training time. For RF model, we find out that RF performances increase by 2.35% to reach 97.83% of accuracy. Same remark for f1-score metric, it increase by 2.33% to reach 97.82%.

TABLE II. KNN, DT and RF ML Models Performances after Applying PCA based on Accuracy and F1-score

	Accuracy	F1-Score
Random Forest	97.83%%	97.83%
Decision Tree	100%	100%
KNN	100%	100%

Table III shows the performances of ML models after selecting the most important features based on RFE feature selection metric. As for PCA metric, the performances of KNN and DT remain the same (100% of accuracy and f1-score for both models). For RF model, accuracy decreased by 0.71% to become 94.77%. Same behaviour for f1-score, it decreased by 0.73% to become 94.76%.

TABLE III. KNN, DT and RF ML Models Performances after Applying RFE based on Accuracy and F1-score

	Accuracy	F1-Score
Random Forest	94.77%	94.76%
Decision Tree	100%	100%
KNN	100%	100%

Fig. 1 shows the confusion matrix of RF model using all features of the dataset. It shows that the accuracy of multi-classification is in the range [94%-97%]. Fig. 2 and Fig. 3 show respectively the multi-classification for both KNN and DT, each class from 0 to 3 are well classified which means that these two models can be a good choice to deploy them as real classifiers for predicting severity of aviation accidents.

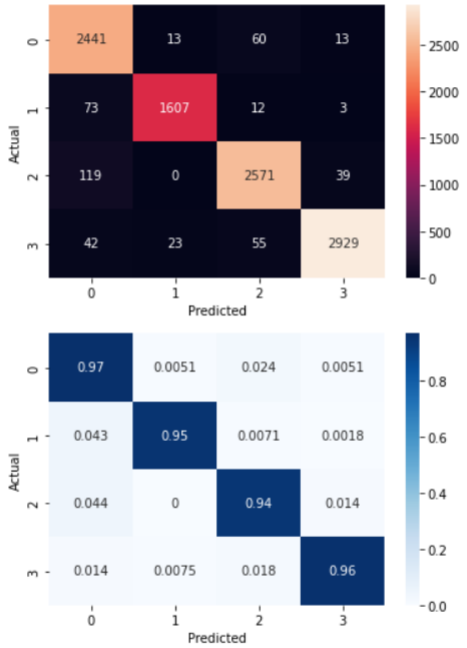


Fig. 1. Multi-classification results of RF model using all features.

Fig. 4 shows the multi-classification performances of RF after application of PCA. We can remark that the range of accuracy for the four classes is [97%-99%]. We also remark that comparing with results using all features, we can say that accuracy augmented for all classes after applying PCA by 2%, 4%, 3% and 1% for respectively class 0, class 1, class 2 and class 3.

Fig. 5 and Fig. 6 show the performance of both KNN and DT. Results remain the same, each class reach 100% of accuracy for the four classes for both KNN and DT; results same the same comparing them with results obtained after application of PCA. Fig. 7 shows the performances of RF model after application of RFE feature selection based method. Results show a decrease of performances by 1%, 1%, 1% for respectively class 0, class 1 and class 2, while the accuracy of class 3 remain the same. Fig. 8 and Fig. 9 show performances of both KNN and DT after application of RFE. Results remain the same such in the two cases (after application of PCA & using all features).

After discussing all these results, we can say that RF,

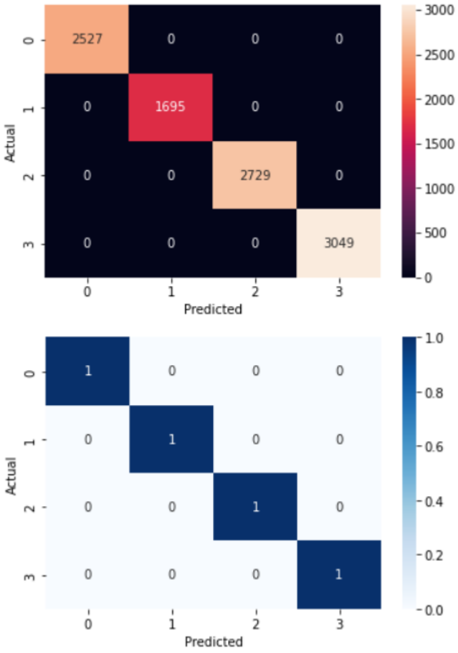


Fig. 2. Multi-classification results of KNN model using all features.

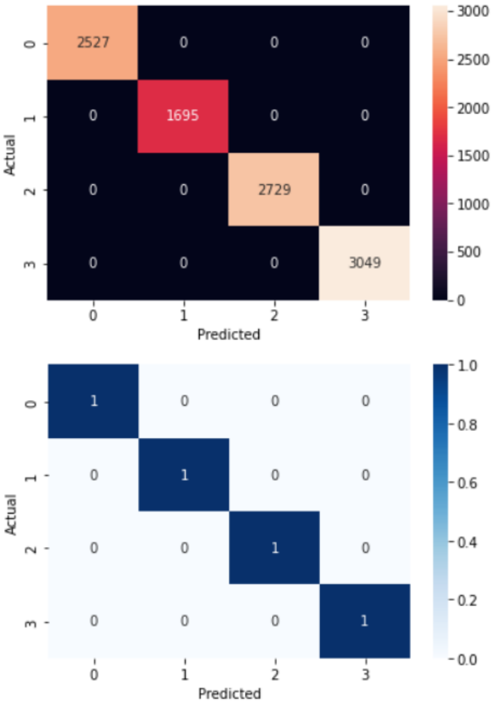


Fig. 3. Multi-classification results of DT model using all features.

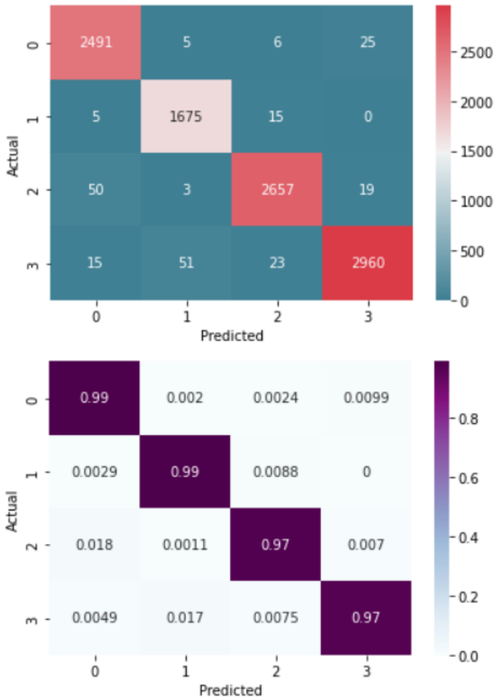


Fig. 4. Multi-classification results of RF model after application of PCA.

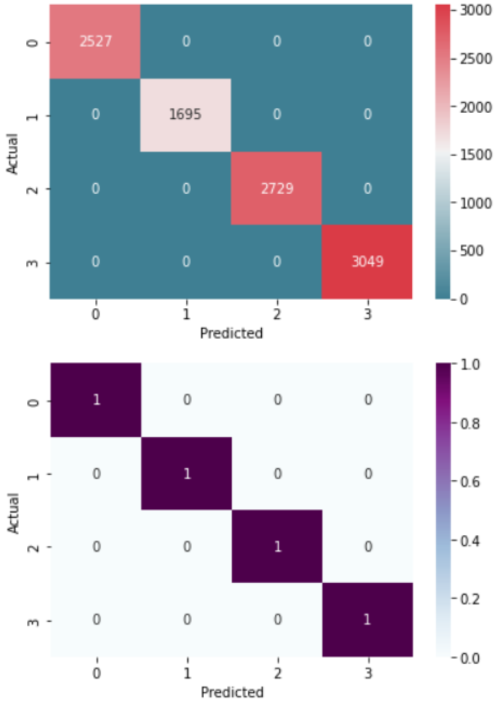


Fig. 5. Multi-classification results of KNN model after application of PCA.

KNN and DT are promising models that can deployed in real situations for predicting severity of aviation accidents. KNN and DT give an accuracy and f1-score equal to 100% which means that all classes of severity have been adequately classified so we can say that intelligent solutions based on ML or Deep Learning models can be relevant alternatives to overcome the limits of classical solutions such as statistical analysis, solutions based on expertise, solutions that require sometimes advanced mathematical modelisation that are complex and may lead sometimes to misrepresent real conditions in the phase of abstraction.

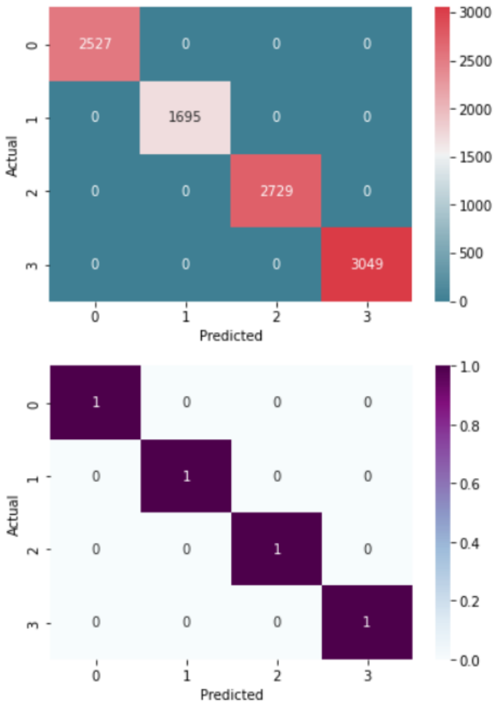


Fig. 6. Multi-classification results of DT model after application of PCA.

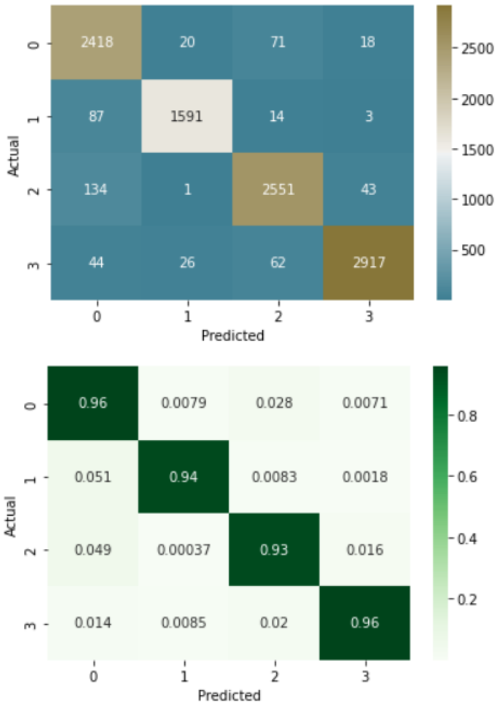


Fig. 7. Multi-classification results of RF model after application of RFE.

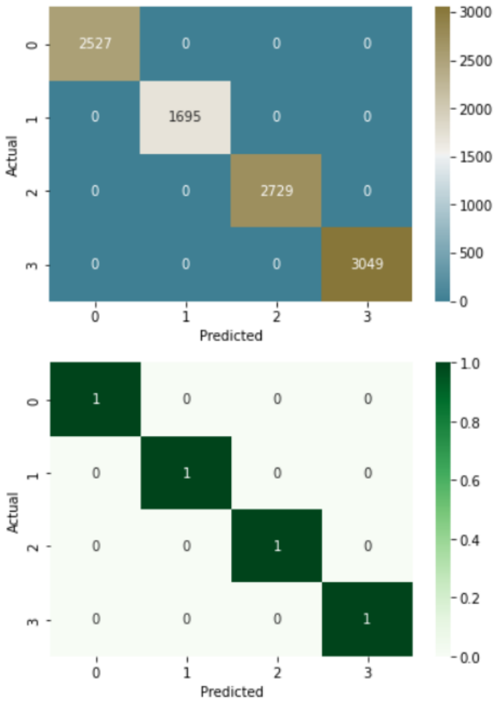


Fig. 8. Multi-classification results of KNN model after application of RFE.

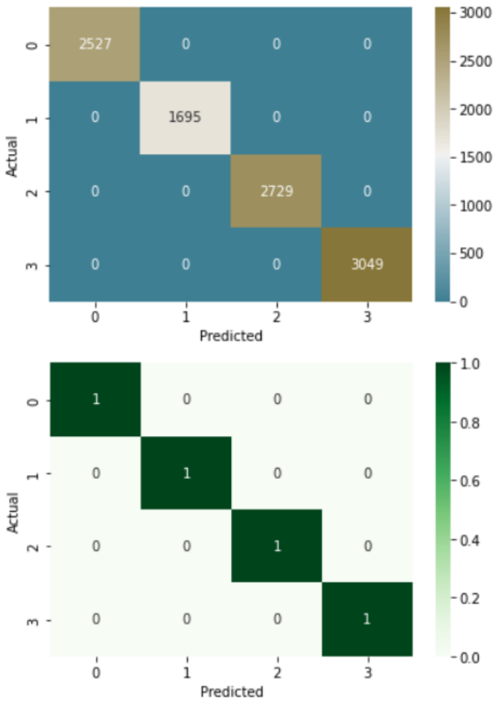


Fig. 9. Multi-classification results of DT model after application of RFE.

V. Conclusion and Future Work

This paper shows the robustness of machine learning models for predicting severity of aviation accidents. KNN and DT achieved high level of performances based on accuracy and f1-score metrics while RF gives respectful results but never achieved 100% of accuracy or f1-score but we should pay attention that these models should be tested in real situations to test its stability. Furthermore, as we had discussed earlier, Neural Network and Deep Learning models can be a good alternatives of the models we had implemented in this study.

This proposition must be the perspective of this data and test our models on other datasets that contain different information and then different conditions. Predictions models in field of aviation are a very complex tasks that require gathering the maximum possible of information about environment plane, human, technologies, etc. in order to improve intelligent solutions like ML and DL models to overcome limits of classical solutions.

REFERENCES

[1] Aviation and Plane Crash Statistics (Updated 2022). (n.d.). Panish — Shea — Boyle — Ravipudi LLP. Retrieved April 8, 2023, from https://www.psbr.law/aviation_accident_statistics.html.

[2] BAUGH, Bradley S. Predicting general aviation accidents using machine learning Algorithms. Embry-Riddle Aeronautical University, 2020.

[3] RAIKAR, Likita J., PARDESHI, Sayali, et SAWALE, Pritam. Airplane Crash Analysis and Prediction using Machine Learning. in International Research Journal of Engineering and Technology (IRJET), vol. 7, no 03,2020.

[4] A.O. Alkhamisi, R. Mehmood, "An ensemble machine and deep learning model for risk prediction in aviation systems", Conf. Data Sci. and Mach. Learn. Appl., Vol. 2020, No. 6, pp. 54-59, Mar. 2020.

[5] H. Kharoufah, J. Murray, G. Baxter, G. Wild, "A review of human factors causations in commercial air transport accidents and incidents: From 2000-2016", Prog. in Aerospace Sci., Vol. 99, pp. 1-13, 2018.

[6] BURNETT, R. Alan et SI, Dong. Prediction of injuries and fatalities in aviation accidents through machine learning. In : Proceedings of the International Conference on Compute and Data Analysis. 2017. p. 60-68.

[7] M. Bazargan, V.S. Guzhva, "Impact of gender, age and experience of pilots on general aviation accidents", Accid. Anal. & Prev., Vol. 43, No. 3, pp. 962-970, 2011.

[8] Kuşkapan, Emre, SAHRAEİ, Mohammad Ali, et Çodur, Muhammed Yasin. Classification of aviation accidents using data mining algorithms. Balkan Journal of Electrical and Computer Engineering, vol. 10, no 1, p. 10-15, 2021.

[9] NIMMAGADDA, SreeRam, SIVAKUMAR, Soubraylu, KUMAR, Naveen, et al. Predicting airline crash due to birds strike using machine learning. In : 2020 7th international conference on smart structures and systems (ICSSS). IEEE. p. 1-4, 2022.

[10] Airplane Accidents Severity Dataset https://www.kaggle.com/datasets/kaushal2896/airplane-accidents-severity-dataset.

[11] Y. Guo, Y. Sun, Y. He, F. Du, S. Su, C. Peng, "A Data-driven Integrated Safety Risk Warning Model based on Deep Learning for Civil Aircraft", IEEE Trans. on Aerospace and Electronic Systems, 2022, pp. 1-14.

[12] ZHANG, Xiaoge et MAHADEVAN, Sankaran. Ensemble machine learning models for aviation incident risk prediction. Decision Support Systems, vol. 116, p. 48-63, 2019.

[13] PANDE, Nikita, GUPTA, Devyani, SHREEMALI, Jitendra and CHAKRABARTI, Prasun. Predicting Fatalities in Air Accidents using CHAID XG Boost Generalized Linear Model Neural Network and Ensemble Models of Machine Learning. Vol. 9 , 30 March 2020.

[14] ZHANG, Xiaoge, SRINIVASAN, Prabhakar, et MAHADEVAN, Sankaran. Sequential deep learning from NTSB reports for aviation safety prognosis. Safety science, vol. 142, p. 105390, 2021.

[15] QUINLAN, J.. Ross . Induction of decision trees. Machine learning, vol. 1, p. 81-106, 1986 .

[16] COSTA, Vinícius G. et PEDREIRA, Carlos E. Recent advances in decision trees: An updated survey. Artificial Intelligence Review, vol. 56, no 5, p. 4765-4800, 2023.

[17] BREIMAN, Leo. Random forests. Machine learning, vol. 45, p. 5-32, 2001.

[18] COVER, Thomas et HART, Peter. Nearest neighbor pattern classification. IEEE transactions on information theory, vol. 13, no 1, p. 21-27, 1967.

[19] HOTELLING, Harold. Analysis of a complex of statistical variables into principal components. Journal of educational psychology, vol. 24, no 6, p. 417, 1933.

[20] GÁRATE-ESCAMILA, Anna Karen, EL HASSANI, Amir Hajjam, et ANDRÈS, Emmanuel. Classification models for heart disease prediction using feature selection and PCA. Informatics in Medicine Unlocked, vol. 19, p. 100330, 2020.

[21] KANNARI, Phanindra Reddy, CHOWDARY, Noorullah Shariff, et BIRADAR, Rajkumar Laxmikanth. An anomaly-based intrusion detection system using recursive feature elimination technique for improved attack detection. Theoretical Computer Science, vol. 931, p. 56-64, 2022.

Enhancing Safety and Multifaceted Preferences to Optimise Cycling Routes for Cyclist-Centric Urban Mobility

Mohammed Alatiyyah
Department of Computer Science, College of Sciences and Humanities-Aflaj,
Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

***Abstract*—In order to optimise bicycle routes across a variety of multiple parameters, including safety, efficiency and subtle rider preferences, this work explores the difficult domain of the Bike Routing Problem (BRP) using a sophisticated Simulated Annealing approach. In this innovative structure, a wide range of limitations and inclinations are combined and carefully calibrated to create routes that skillfully meet the varied and changing needs of cyclists. Extensive testing on a dataset representing a range of rider preferences demonstrates the effectiveness of this novel approach, resulting in significant improvements in route selection. This research is a significant resource for urban planners and politicians. Its data-driven solutions and strategic recommendations will help them strengthen bicycle infrastructure, even beyond its immediate applicability in resolving the BRP.**

***Keywords*—Bike routing; dynamic vehicle routing inventory routing; approximate dynamic programming**

I. Introduction

When compared to driving, biking is an affordable, environmentally friendly, energy-efficient, and health-conscious alternative to other forms of transportation [1]. It goes beyond simple transportation, providing a means of regular physical activity, encouraging wholesome living, and reducing car emissions. Even with the growing number of towns attempting to develop networks of bicycle lanes, a significant percentage of trips within rideable distances are still made by automobile [2]. Promoting a mentality that embraces riding for recreational and everyday transportation purposes is still a critical obstacle to creating genuinely bike-friendly urban settings.

The key to solving this problem is to carefully hone the bike-path network's functionality and architecture. Riding a bicycle is one of the most effective ways to use energy and promote public health. It is also one of the most important forms of active transportation to reduce traffic congestion and avoid emissions from vehicles. Because walking and cycling are accessible and don't require any specific equipment or expertise, they are suitable for people of all ages and allow them to customise the amount of physical effort they want to put in. But if walking is more suitable for shorter distances, cycling is a better choice for longer ones. However, the belief that riding a bicycle is a dangerous activity endures, mostly because of things like heavy traffic, congested roads, and a lack of designated bike lanes. There are several obstacles that prevent cycling from becoming widely accepted, including worries about comfort, safety, and accessibility [3]. Safety barriers are significant difficulties that arise from concerns about criminal activity, road accidents, or personal injury. Cycling enthusiasts frequently view themselves as vulnerable users of a space designed primarily for motor vehicles. The selection of bike routes is a significantly more complex procedure than the selection of driving routes. Whereas motorists focus on distance and travel time [4], cyclists consider a variety of factors, such as the condition of the bike lanes that are available [5][6] and avoiding hilly terrain and particularly unsettling intersections like roundabouts [7]. Bicyclists' preferred routes are greatly influenced by their closeness to motorised vehicles, especially on high-speed main roads [6]. In order to avoid heavily trafficked or densely populated bike regions, cyclists frequently choose longer but supposedly safer routes [8][9]. This paper introduces a novel approach to the Bike Route Problem (BRP), with a focus on cycling route optimisation to improve safety, efficiency, and compatibility with a range of rider preferences. The new framework synthesises a wide range of limitations and preferences, allowing bicycle routes to be customised to meet the diverse needs of different riders. Our study provides empirical confirmation of the suggested method's effectiveness using a dataset that encompasses a wide range of rider preferences. The flexibility of the Simulated Annealing method demonstrates its ability to create customised routes that meet the various needs of cyclists.

The Bike Route Problem is explained in Section III, while the following sections outline the state of relevant research in Section II. Section IV presents our suggested options for solving the problem, followed by an in-depth analysis of the findings. In the end, Section V summarizes the findings and suggests possible avenues for more research and development.

II. Related Work

Bicycle routing is a problem that has been extensively studied in a number of research paradigms, both directly and indirectly. The multi-objective routing problem was first tackled by Martins et al. [10], who defined the problem in terms of several objective functions and developed a multi-criteria label-setting algorithm for its solution. Later research, like that of Song et al. [11], explored the use of multi-label correction algorithms to find Pareto routes and included hierarchical clustering techniques to expedite the selection procedure. Even with recent improvements in label-setting algorithms, real-time applications are still hindered by the processing timescales these techniques demand.

Routing difficulties have seen the use of evolutionary approaches, most notably genetic algorithms. Genetic algorithms were used in conjunction with High-performance Clusters (HPC) by Arunadevi et al. [12] to address routing issues. Comparably, Kang et al. [13] used evolutionary algorithms to compute segment-specific cost functions in networks of cyclists, taking perceived risk and distance into account. Nevertheless, the multi-objective landscape of bicycle routing was not explored in these research, which mostly concentrated on single-objective optimisation. Several studies have used various criteria to build bicycle routes, frequently maximising each criterion on its own. Using ArcGIS Server and specific Google APIs, Hochmair et al. [14] developed an online cycling route planner. Hrncir et al. [15] used a cost vector that included several parameters such as trip time, comfort, quietness, and levelness with the A-Star algorithm. Chen et al. [16] investigated the use of artificial neural networks to generate routing algorithm heuristic functions. Contraction hierarchies were combined with OpenStreetMap data by Luxen et al. [17] to determine the most straightforward graph-based pathways. Methods for solving the multi-objective bike routing problem have been developed recently, and their effectiveness has been demonstrated in real-world network scenarios. Bike routing algorithms were improved by Hrncir et al. [18] and Hrnvcivr et al. [19], who presented a heuristic-driven Dijkstra algorithm that included a multi-criteria viewpoint. Notable platforms have surfaced in the world of bicycle-specific internet route planners. Though it is available in some areas, Google Maps does not provide much personalization. Using routing techniques like the A* algorithm and contraction hierarchies, OpenTripPlanner incorporates a bicycle planning option that allows users to balance preferences like speed and terrain flatness. Popular in the United Kingdom, Cyclestreets offers a range of route alternatives depending on balance, speed, and tranquilly. Berlin, Germany-based BBBike takes into account variables including the kind of route and the presence of lights. Still, there isn't much written about this topic in the literature. A few studies—Robert et al. [20] and Su et al. [21], for example—introduced computerised cycling route planners customised for certain areas, while Hochmair et al. [14] offered a bicycle route planner for Broward County, Florida, taking into account a variety of factors influenced by cyclist preferences. A route planner for electric bicycles was presented by Tal et al. [22], with an emphasis on weather and energy efficiency.

Although bike routing has progressed, there is still a significant gap in the field of multi-objective methods that balance optimising complicated objectives with computing feasibility. In order to close this gap, this work presents a novel approach that uses genetic algorithms to approximate the optimal Pareto set. It also investigates the possibility of using genetic algorithms to solve multi-objective problems in a reasonable amount of computational time.

III. Bike Routing Problem (BRP)

We define the BRP as follows: let $G = \{V, E\}$ be a undirected weighted graph where, $i \in V$ and $i = 1, 2, \dots, |V|$ be a set of nodes, and E be a set of edges between nodes where E_{ij} be the edge between V_i and V_j. In addition, each edge has a set attributes denotes A where $a \in A$ and $a = 1, 2, \dots, |A|$. The starting node is s and the terminal node is e where $s = 1$ and $e = |V|$. However, cyclists have number of constraints. Mandatory Constraints (MC) be denoted MC; $mc \in MC$, where $mc = 1, 2, \dots, |MC|$. Optional Constraints (OC) be denoted OC; $oc \in OC$, where $oc = 1, 2, \dots, |OC|$. The total number of constraints $|MC| + |OC| \leq |A|$. In addition, each constraints (MC or OC) be applied in each edge and be denoted MC_{ij}^{mc} or OC_{ij}^{oc} where MC or $OC \subseteq A$.

$$Max \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} A_{ij} \times X_{ij} \tag{1}$$

$$\Big(|MC| + |OC|\Big) \leq |A| \tag{2}$$

$$\sum_{i=1}^{|V|} X_{si} = 1 \tag{3}$$

$$\sum_{i=1}^{|V|} X_{ie} = 1 \tag{4}$$

$$X_{ij} \in \{1, 0\}; \forall i, j = 1, 2, \dots, |V| \tag{5}$$

$$\sum_{n=1}^{|V|-1} X_{nr} = \sum_{n=2}^{|V|} X_{ru} = 1 \tag{6}$$

$\forall r = 2, \dots, |V| - 1$

$$2 \leq I_n \leq |V| \tag{7}$$

$\forall n = 2, \dots, |V|$

$$I_n - I_u + 1 \leq (|V| - 1) \times (1 - X_{nu}) \tag{8}$$

$\forall n, u = 2, \dots, |V|$

Eq. 1 presents the objective function of BRP where X_{ij} denotes the decision variable moving from node i to node j and the the value of X_{ij} is 0 or 1 (see Eq. (5)). Eq. 3 and 4 represent a constraint to ensure the path starts from s and ends at e. Eq. 6 is a constraint to ensure that the path is connected and each vertex is visited once at most. Eq. 7, I_n denotes the position of node n in the path, and the combination of Eq. 7 and Eq. 8 prevents sub roues.

$$MC_{ij} = \prod_{mc=1}^{|MC|} MC_{ij}^{mc} \tag{9}$$

$$_{ij}^{mc} \in \{1, 0\} \tag{10}$$

$$\forall i, j = 1, 2, \dots, |V| \text{ and } \forall mc = 1, 2, \dots, |MC|$$

As have been mentioned above that there are numbers of MCs which is donated in Eq. 9. Mainly, MC_{ij}^{mc} denotes the MC (mc) from node i and j where MC_{ij}^{mc} has one value if the constraint is satisfied, the value is equal 1 otherwise equal 0 (see Eq. (10)).

TABLE I. INSTANCE TEST SCENARIO CATEGORIES

Scenario	Description	Details
S1	Commuter Cyclist	Using bike for work commuting
S2	Fitness Enthusiast	Cycling for exercise and fitness
S3	Urban Explorer	Exploring the city and its surroundings
S4	Nature Lover	Cycling in natural landscapes and parks
S5	Daily Commuter	Regular commuting for work and daily activities
S6	Adventurous Cyclist	Exploring challenging terrains and trails
S7	Family Outings	Cycling with family for recreational activities
S8	Bike Commuters	Using bike for work commuting in a busy city
S9	Night Rider	Cycling during nighttime for relaxation
S10	Touring Cyclist	Long-distance touring and exploration

TABLE II. INSTANCE TEST SCENARIO IN MORE DETAILS

Scenario	Attributes				
	Safety	Bike Lanes	Traffic Volume	Scenery	Elevation
S1	MC	OC	OC (M)	OC	OC (Low)
S2	OC	OC	OC (L)	MC	MC
S3	OC	OC	OC (M)	MC	OC (Low)
S4	MC	OC	OC (L)	MC	OC (Low)
S5	MC	MC	OC (L)	OC	OC (Low)
S6	OC	OC	OC (L)	OC	OC (High)
S7	MC	OC	OC (L)	OC	OC (Low)
S8	OC	OC	MC	OC	OC (Low)
S9	OC	OC	OC (L)	OC	OC (Low)
S10	MC	OC	OC (L)	OC	OC (High)

$$OC_{ij} = \frac{\sum_{oc=1}^{|OC|} OC_{ij}^{oc}}{|OC|} \tag{11}$$

$$0 \leq OC_{ij}^{oc} \leq 1 \tag{12}$$

$$\forall i, j = 1, 2, \ldots, |N| \text{ and } \forall oc = 1, 2, \ldots, |OC|$$

In contrast, a OC indicates a specific level of satisfaction and meeting it is optional. Numbers of OCs which is donated in Eq. 11, where OC_{ij}^{oc} one value between 0 to 1 (see Eq. (12)). In additional, Eq. 13 presents the calculation of MC and OC from i to j.

$$A_{ij} = MC_{ij} \times OC_{ij} \tag{13}$$

IV. SOLUTION APPROACHES

In this section, we elaborate on our heuristic-driven Simulated Annealing method designed for solving the BRP. Our approach integrates a data model for constraints and Simulated Annealing techniques to address the Bike Route Problem. The fundamental concept involves simplifying the complexity of the problem by consolidating constraints into a single value that encapsulates the attributes sought by riders.

A. Benchmark Instances

To the best author knowledge there is not any dataset for bike routing problem, so A set of benchmark instances were created to analyze how the propose model performs through numerical experiment results. Random problem instances were generated so as to maintain the properties of one of ten general scenario categories as defined in Table I. In each scenario, it has been generated different circumstances where difference constraints and preferences are applied. Each instance was randomly generated assuming a grid of 40 by 40 miles based on an area similar in size Newcastle upon Tyne, UK. Table II shows the details of each scenario, and in the dataset has been created five attributes for each of edges.

Table II delineates a variety of constraints and preferences. The Optional Constraints (OC) column details the preferences of the riders, while the Mandatory Constraints (MC) column lists the constraints that are requisite. Furthermore, the Traffic Volume attribute accommodates varying preferences: certain riders opt for routes with low traffic, whereas others may favor routes with moderate traffic levels. Additionally, the Elevation attribute reflects diverse inclinations regarding physical exertion; some riders seek routes with minimal elevation to reduce effort, while others pursue routes with significant elevation for a more challenging ride.

TABLE III. INSTANCE SCENARIO WITH START AND END LOCATION

Instance	Start location	End location
I1	121	165
I2	77	351
I3	462	145
I4	282	393
I5	25	115
I6	80	476
I7	323	342
I8	109	464
I9	491	171
I10	31	375

The ten benchmark sets under consideration encompass a total of 100 instances, with each set comprising 10 instances characterized by a node count of 500. While each benchmark (scenario) shares identical start and end points, they are differentiated by their unique constraints and preferences. Table III details the start and end points for each instance, providing a clear reference for the scenarios tested.

Algorithm 1 Bike Routing Problem

```
$SenarioData \leftarrow ReadingDataFiles()$
$Edges \leftarrow ReadingEdges()$
while $i <= SenarioData.lenth()$ do
    while $j <= Edges.lenth()$ do
        $InitialRoute \leftarrow CreateInitialRoute()$
        $InitialTemperature = 1000$
        $CoolingRate = 0.995$
        $BestRoute = SimulatedAnnealing()$
        $j \leftarrow j + 1$
    end while
    $i \leftarrow i + 1$
end while
```

B. Simulated Annealing

Simulated Annealing (SA) is a powerful optimization algorithm inspired by the annealing process in metallurgy. Initially introduced by Kirkpatrick, Gelatt, and Vecchi in the 1980s [23], SA mimics the annealing of materials, where a solid is heated to high temperatures and then gradually cooled to minimize its energy state. This process allows the algorithm to escape local optima and explore the solution space more effectively. The key idea behind SA is to accept worse solutions

Algorithm 2 $SimulatedAnnealing$ Algorithm

```
while $temperature > 0.1$ do
    while $ReplacedNode$ do
        $SwapIndex \leftarrow random(1, Route.length())$
        $NewNode \leftarrow FindReplacedNode()$
        if $NewNode! = Null$ then
            $Route[SwapIndex] \leftarrow NewNode$
            $Update =$
        else
            $ReplacedNode \leftarrow False$
        end if
    end while
end while
```

TABLE IV. The Result for Instance 1

scenario	Total Scores	Path
Scenario-1	[121, 297, 422, 165]	0.92
Scenario-2	[121, 297, 346, 165]	0.81
Scenario-3	[121, 297, 346, 165]	0.68
Scenario-4	[121, 319, 309, 165]	0.82
Scenario-5	[121, 297, 422, 165]	1.00
Scenario-6	[121, 319, 309, 165]	0.33
Scenario-7	[121, 297, 143, 165]	0.59
Scenario-8	[121, 297, 422, 165]	0.84
Scenario-9	[121, 239, 257, 249, 165]	0.35
Scenario-10	[121, 297, 143, 165]	0.51

TABLE V. The Result for Instance 2

Scenario	Total Scores	Path
Scenario-1	[77, 422, 351]	0.33
Scenario-2	[77, 422, 351]	0.78
Scenario-3	[77, 422, 351]	0.33
Scenario-4	[77, 422, 351]	0.50
Scenario-5	[77, 422, 351]	0.55
Scenario-6	[77, 422, 351]	0.28
Scenario-7	[77, 422, 351]	0.35
Scenario-8	[77, 422, 351]	1.00
Scenario-9	[77, 422, 351]	0.20
Scenario-10	[77, 422, 351]	0.43

TABLE VI. The Result for Instance 3

Instance	Total Scores	Path
Scenario-1	[462, 86, 145]	0.83
Scenario-2	[462, 86, 145]	1.00
Scenario-3	[462, 41, 2, 145]	0.63
Scenario-4	[462, 86, 145]	1.00
Scenario-5	[462, 86, 145]	0.93
Scenario-6	[462, 86, 145]	0.46
Scenario-7	[462, 86, 145]	0.83
Scenario-8	[462, 86, 145]	0.61
Scenario-9	[462, 86, 145]	0.56
Scenario-10	[462, 86, 145]	0.76

TABLE VII. The Result for Instance 4

Instance	Total Scores	Path
Scenario-1	[282, 291, 148, 393]	0.57
Scenario-2	[282, 291, 243, 393]	0.83
Scenario-3	[282, 366, 144, 393]	0.50
Scenario-4	[282, 366, 144, 393]	0.60
Scenario-5	[282, 291, 243, 393]	0.61
Scenario-6	[282, 455, 325, 393]	0.26
Scenario-7	[282, 291, 243, 393]	0.41
Scenario-8	[282, 291, 243, 393]	0.82
Scenario-9	[282, 455, 325, 393]	0.23
Scenario-10	[282, 291, 243, 393]	0.49

with a certain probability, enabling the algorithm to explore the solution space broadly before converging towards the optimal solution [23][24]. This approach has proven to be highly effective in solving complex optimization problems where the objective function is not explicitly defined and can only be evaluated through computationally expensive simulations [24]. SA's ability to balance exploration and exploitation makes it a popular choice in various real-world applications, ranging from engineering and logistics to machine learning and artificial intelligence [24]. Its widespread applicability and efficiency in tackling challenging optimization problems have solidified its position as a prominent metaheuristic algorithm in the field of computational optimization.

V. Computational Results

In this section, we provide the outcomes of our numerical experiments. Initially, we assess the efficiency of our BRP formulation as well as the Constraints formulation coupled with Simulated Annealing. Subsequently, we analyze the performance of our proposed Simulated Annealing algorithm. All experiments were carried out on a computer with an Intel i7-2.10 GHz processor and 64 GB RAM, running on the Windows 11-x64 operating system.

The study examined the effectiveness of BRP in uneven situations. Results from the tenth series of tests were analyzed in comparison with outcomes obtained using Simulated Annealing methods. Tables IV - XIII summarize the computational results for each instance, respectively. In these tables, the three columns display the scenario name, total scores based on the edges are visited, and the path. The total scores values are evaluated based in Eq. 1. Please note that, the computation times of these heuristics are less than 1s.

Table IV reports the results obtained from the test problem in instance-1 which representing a scenario of transporting from Node-121 to Node-165. As it can be seen, the results can be categorized into five categories: (1) Scenario 1, 5, and 8, (2) Scenario 2 and 3, (3) Scenario 4 and 6, (4) Scenario 7 and 10, and (5) Scenario 9; each one of these category has the same result.

Table V presents the results of the instance-2 which presents the problem of moving from Node-77 to Node-351. In addition, Table XII presents the results of the instance-9 which presents the problem of moving from Node-491 to Node-171. Surprisingly, all results from different scenarios has the same result (path); In instance-2 the result is [77, 422, 351], and the instance-9 the result is [491, 218, 171].

Table VI shows the results of the instance-3 which presents the problem of moving from Node-462 to Node-145. In this experiment, All scenarios shows the same results (path) except the scenario-3.

Table VII presents the results of the instance-3 which presents the problem of moving from Node-282 to Node-393. The results can be seen in four categories: (1) Scenario 2, 5, 7, 8, and 10, (2) Scenario 3 and 4, (3) Scenario 6 and 9, and (4) Scenario 1.

Table VIII presents the results of the instance-2 which presents the problem of moving from Node-25 to Node-115. There are six scenarios have different result completely which

TABLE VIII. The Result for Instance 5

Instance	Total Scores	Path
Scenario-1	[25, 240, 184, 115]	0.56
Scenario-2	[25, 18, 108, 115]	1.00
Scenario-3	[25, 497, 131, 115]	0.47
Scenario-4	[25, 117, 285, 115]	0.72
Scenario-5	[25, 18, 108, 115]	0.95
Scenario-6	[25, 383, 46, 115]	0.49
Scenario-7	[25, 75, 110, 115]	0.60
Scenario-8	[25, 240, 184, 115]	0.70
Scenario-9	[25, 462, 378, 285, 115]	0.41
Scenario-10	[25, 281, 170, 454, 115]	0.49

TABLE IX. The Result for Instance 6

Instance	Total Scores	Path
Scenario-1	[80, 404, 476]	0.74
Scenario-2	[80, 404, 476]	0.91
Scenario-3	[80, 404, 476]	0.52
Scenario-4	[80, 404, 476]	0.83
Scenario-5	[80, 404, 476]	0.87
Scenario-6	[80, 404, 476]	0.42
Scenario-7	[80, 40, 166, 476]	0.46
Scenario-8	[80, 404, 476]	1.00
Scenario-9	[80, 404, 476]	0.42
Scenario-10	[80, 404, 476]	0.70

TABLE X. The Result for Instance 7

Instance	Total Scores	Path
Scenario-1	[323, 421, 342]	0.56
Scenario-2	[323, 421, 342]	0.67
Scenario-3	[323, 421, 342]	0.56
Scenario-4	[323, 421, 342]	0.61
Scenario-5	[323, 421, 342]	0.67
Scenario-6	[323, 437, 313, 342]	0.38
Scenario-7	[323, 421, 342]	0.58
Scenario-8	[323, 421, 342]	0.67
Scenario-9	[323, 421, 342]	0.31
Scenario-10	[323, 421, 342]	0.53

TABLE XI. The Result for Instance 8

Instance	Total Scores	Path
Scenario-1	[109, 18, 100, 464]	0.68
Scenario-2	[109, 106, 498, 464]	0.77
Scenario-3	[109, 125, 377, 464]	0.49
Scenario-4	[109, 370, 243, 464]	0.83
Scenario-5	[109, 225, 214, 464]	0.90
Scenario-6	[109, 341, 498, 464]	0.32
Scenario-7	[109, 432, 377, 464]	0.67
Scenario-8	[109, 370, 202, 464]	0.64
Scenario-9	[109, 20, 458, 268, 464]	0.38
Scenario-10	[109, 252, 56, 464]	0.79

TABLE XII. The Result for Instance 9

Instance	Total Scores	Path
Scenario-1	[491, 218, 171]	0.64
Scenario-2	[491, 218, 171]	0.92
Scenario-3	[491, 218, 171]	0.84
Scenario-4	[491, 218, 171]	0.90
Scenario-5	[491, 218, 171]	1.00
Scenario-6	[491, 218, 171]	0.59
Scenario-7	[491, 218, 171]	0.63
Scenario-8	[491, 218, 171]	0.80
Scenario-9	[491, 218, 171]	0.47
Scenario-10	[491, 218, 171]	0.75

TABLE XIII. The Result for Instance 10

Instance	Total Scores	Path
Scenario-1	[31, 356, 375]	1.00
Scenario-2	[31, 356, 375]	0.95
Scenario-3	[31, 356, 375]	0.88
Scenario-4	[31, 356, 375]	1.00
Scenario-5	[31, 336, 26, 375]	0.60
Scenario-6	[31, 356, 375]	0.62
Scenario-7	[31, 356, 375]	1.00
Scenario-8	[31, 356, 375]	0.63
Scenario-9	[31, 336, 26, 375]	0.24
Scenario-10	[31, 356, 375]	1.00

are Scenario 3, 4, 6, 7, 9, and 10. However, Scenarios 1 and 8 are the same, and the Scenario 2 and 5 are the same.

Table IX shows the results of the instance-6 which presents the problem of moving from Node-80 to Node-476. In this experiment, All scenarios shows the same results (path) except the scenario-7. Moreover, Table X shows the results of the instance-7 which presents the problem of moving from Node-323 to Node-342. Also, ins this experiment, All scenarios shows the same results (path) except the scenario-6.

Table XI shows the results of the instance-8 which presents the problem of moving from Node-109 to Node-464. Surprisingly, all results are different for each scenario.

Table XIII presents the results of the instance-10 which presents the problem of moving from Node-31 to Node-375. The results can be seen in two categories: (1) Scenario 1, 2, 3, 4, 6, 7, 8 and 10, and (2) Scenario 5 and 9.

VI. Conclusions and Future Work

The study successfully applied a heuristic-driven Simulated Annealing algorithm to the BRP, demonstrating its efficacy in processing and optimizing complex routing problems within reasonable computational times. The results confirmed that the proposed method could handle a variety of scenarios by accommodating diverse constraints and preferences, thus offering a flexible and robust solution to the BRP. The findings suggest that the method is not only applicable in the context of cycling but may also extend to other forms of transportation where route optimization is essential. The research contributes to the field by providing a systematic approach to addressing BRP and paving the way for more sustainable urban transport systems. Future research directions include scaling the proposed solution to larger datasets and urban areas with more complex networks. There is also scope for integrating real-time data, such as traffic updates and weather conditions, to enhance the dynamicity and responsiveness of the route planning process. Another avenue for exploration is the application of the Simulated Annealing approach to different types of multi-objective routing problems beyond cycling, such as pedestrian pathfinding and electric vehicle charging station routes. Further studies could also investigate the integration of machine learning techniques to predict and adapt to cyclists' preferences more accurately.

References

[1] S. Ryu, A. Chen, J. Su, and K. Choi, “Two-stage bicycle traffic assignment model,” *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 2, p. 04017079, 2018.

[2] B. E. Saelens, J. F. Sallis, and L. D. Frank, “Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures,” *Annals of behavioral medicine*, vol. 25, no. 2, pp. 80–91, 2003.

[3] D. Piatkowski, R. Bronson, W. Marshall, and K. J. Krizek, “Measuring the impacts of bike-to-work day events and identifying barriers to increased commuter cycling,” *Journal of Urban Planning and Development*, vol. 141, no. 4, p. 04014034, 2015.

[4] J. P. Schmitt and F. Baldo, “A method to suggest alternative routes based on analysis of automobiles’ trajectories,” in *2018 XLIV Latin American Computer Conference (CLEI)*. IEEE, 2018, pp. 436–444.

[5] J. Pucher and R. Buehler, “Making cycling irresistible: lessons from the netherlands, denmark and germany,” *Transport reviews*, vol. 28, no. 4, pp. 495–528, 2008.

[6] R. Buehler and J. Dill, “Bikeway networks: A review of effects on cycling,” *Transport reviews*, vol. 36, no. 1, pp. 9–27, 2016.

[7] M. Rasanen and H. Summala, “Car drivers’ adjustments to cyclists at roundabouts,” *Transportation Human Factors*, vol. 2, no. 1, pp. 1–17, 2000.

[8] N. Y. Tilahun, D. M. Levinson, and K. J. Krizek, “Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey,” *Transportation Research Part A: Policy and Practice*, vol. 41, no. 4, pp. 287–301, 2007.

[9] S. E. Vedel, J. B. Jacobsen, and H. Skov-Petersen, “Bicyclists’ preferences for route characteristics and crowding in copenhagen–a choice experiment study of commuters,” *Transportation Research Part A: Policy and Practice*, vol. 100, pp. 53–64, 2017.

[10] E. Q. V. Martins, “On a multicriteria shortest path problem,” *European Journal of Operational Research*, vol. 16, no. 2, pp. 236–245, 1984.

[11] Q. Song, P. Zilecky, M. Jakob, and J. Hrncir, “Exploring pareto routes in multi-criteria urban bicycle routing,” in *17th international IEEE conference on intelligent transportation systems (ITSC)*. IEEE, 2014, pp. 1781–1787.

[12] J. Arunadevi, A. Johnsanjeevkumar, and N. Sujatha, “Intelligent transport route planning using parallel genetic algorithms and mpi in high performance computing cluster,” in *15th International Conference on Advanced Computing and Communications (ADCOM 2007)*. IEEE, 2007, pp. 578–583.

[13] L. Kang and J. D. Fricker, “Bicycle-route choice model incorporating distance and perceived risk,” *Journal of Urban Planning and Development*, vol. 144, no. 4, p. 04018041, 2018.

[14] H. Hochmair and F. Zhaohui, “Web based bicycle trip planning for broward county, florida. gis center,” 2013.

[15] J. Hrncir, Q. Song, P. Zilecky, M. Nemet, and M. Jakob, “Bicycle route planning with route choice preferences,” in *ECAI 2014*. IOS Press, 2014, pp. 1149–1154.

[16] H.-C. Chen and J.-D. Wei, “Using neural networks for evaluation in heuristic search algorithm,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 25, no. 1, 2011, pp. 1768–1769.

[17] D. Luxen and C. Vetter, “Real-time routing with openstreetmap data,” in *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2011, pp. 513–516.

[18] J. Hrncir, P. Zilecky, Q. Song, and M. Jakob, “Speedups for multi-criteria urban bicycle routing,” in *15th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2015)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

[19] J. Hrnčíř, P. Žileckỳ, Q. Song, and M. Jakob, “Practical multicriteria urban bicycle routing,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 493–504, 2016.

[20] R. J. Turverey, D. D. Cheng, O. N. Blair, J. T. Roth, G. M. Lamp, and R. Cogill, “Charlottesville bike route planner,” in *2010 IEEE Systems and Information Engineering Design Symposium*. IEEE, 2010, pp. 68–72.

[21] J. G. Su, M. Winters, M. Nunes, and M. Brauer, “Designing a route planner to facilitate and promote cycling in metro vancouver, canada,” *Transportation research part A: policy and practice*, vol. 44, no. 7, pp. 495–505, 2010.

[22] I. Tal, A. Olaru, and G.-M. Muntean, “ewarpe-energy-efficient weather-aware route planner for electric bicycles,” in *2013 21st IEEE International Conference on Network Protocols (ICNP)*. IEEE, 2013, pp. 1–6.

[23] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, “Optimization by simulated annealing,” *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[24] L. Ingber, “Adaptive simulated annealing (asa): Lessons learned,” *arXiv preprint cs/0001018*, 2000.

Recognition and Translation of Ancient South Arabian Musnad Inscriptions

Afnan Altalhi, Atheer Alwethinani, Bashaer Alghamdi, Jumanah Mutahhar, Wojood Almatrafi, Seereen Noorwali
College of Computers and Information System, Umm AlQura University
Makkah, Saudi Arabia

***Abstract*—Inscriptions play an important role in preserving historical information. As such, conservation of these inscriptions provides valuable insights into the history and cultural heritage of the region. *Musnad* inscriptions are considered one of the earliest forms of writing from the Arabian Peninsula, preceding the modern Arabic font; however, most Musnad inscriptions remain unread and untranslated, signifying a substantial loss of historical information. In response, this paper represents a significant contribution to the field by proposing a successful approach to interpreting Musnad inscriptions. To do so, a dataset was prepared from the Saudi Arabian Ministry of Culture and subjected to preprocessing for optimal recognition, a step that entailed several experiments to enhance image quality and preparedness for recognition. The dataset was then trained and tested with 29 classes using three different convolutional neural network (CNN) architectures: Visual Geometry Group 16 (VGG16), Residual Network 50 (ResNet50) and MobileNetV2. Thereafter, the performance of each architecture was evaluated based on its accuracy in recognising Musnad inscriptions. The results demonstrate that VGG16 achieved the highest accuracy of 93.81%, followed by ResNet50 at 89.39% and MobileNetV2 at 80.02%.**

***Keywords*—Musnad inscriptions; text recognition; deep learning; VGG16; ResNet-50; MobileNetV2**

I. Introduction

The ancient writings and inscriptions found in the Arabian Peninsula hold great significance in the modern era, as they serve as a source for historians and researchers studying ancient history and civilisations. Within the Kingdom of Saudi Arabia, numerous archaeological sites are covered with inscriptions and ancient texts written in the form of Musnad inscriptions, which represent a rich cultural heritage and hold great historical value.

The Musnad inscriptions, originating in the southern region of the Arabian Peninsula, predates the current Arabic font, with Musnad inscriptions discovered on diverse surfaces, from forts and mountainsides to smaller stone fragments and statue bases [1]. These inscriptions provide invaluable insights into the lives of the individuals who created them, including their lifestyle, beliefs, political climate and relationships with neighbouring nations. In addition, Musnad inscriptions were utilised in daily interactions, further highlighting their importance [2] [3].

The Musnad inscriptions is comprised of 29 characters, and the writing direction is right to left. Unlike Arabic font, the characters in Musnad are separate and unconnected in a word. This means the shape of each character remains the same, regardless of its position in the word. As well, words are separated by a vertical line (|). Furthermore, Musnad does not include any punctuation or diacritical marks [2].

In Fig. 1, an example of a Musnad inscription from Al-Faw village is depicted. To translate these inscriptions, each Musnad character is converted into its corresponding counterpart in Standard Arabic, as shown in Fig. 2, where the translation considers the vertical line that separates each word. It is important to note that the translated words belong to an old Arabic dialect, but they are all Arabic [4].

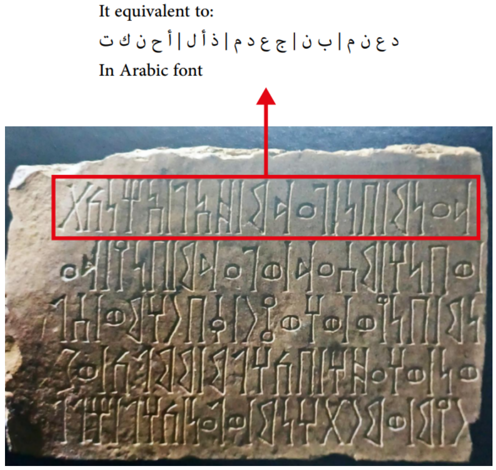


Fig. 1. One of the musnad inscriptions in the village of Al-Faw.

Motivated by the limitations of existing translation methods, such as the time-consuming nature of manual techniques and the dependence on expert availability, this research presents the first attempt to automate the recognition and translation of Musnad inscriptions into Arabic using image processing and deep learning. Consequently, it seeks to enhance the experience of reading Musnad inscriptions, making them more accessible and effortless to understand. Thus, the contributions of this work can be summarised as follows:

- Creation of a dataset for the Musnad inscriptions.
- Application of an optimal image processing technique to improve recognition accuracy.
- Comparison of the performances of three deep learning models (ResNet50, MobileNetv2 and VGG16) to determine which has the highest accuracy.

The remainder of the paper is structured as follows: Section II presents a literature review, Section III provides a detailed description of the Musnad inscription dataset, Section IV outlines the proposed methodology, Section V presents the

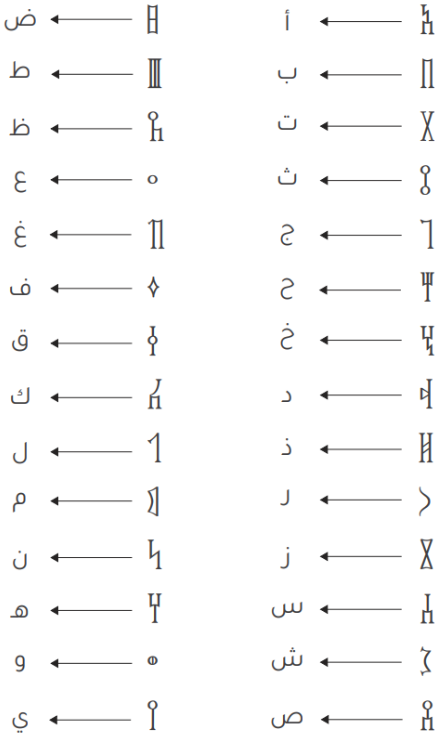


Fig. 2. Musnad alphabet.

results and discussions and Section VI concludes the paper and suggests future directions for research.

II. Literature Review

Scientific research has adopted different approaches to implementing detection and recognition, depending on the inscriptions, ancient characters and language itself. This section will thus provide an overview of related works in our field.

According to previous research, you only look once (YOLO) object detection methods are used in [5], [6], which implements the CNN architecture. YOLOv3-tiny was able to recognise the Kawi character on copper inscriptions due to its high detection accuracy (average of 97.93% in [5] and high detection speed. Meanwhile, the Oracle Bone inscriptions (OBIs) were recognised using two deep learning models in [6]: first, YOLOv3-tiny was used to detect and recognise OBIs, and second, MobileNet was used to detect undetected OBIs, as YOLOv3-tiny's limitations prevent all OBIs from being correctly recognised. Thus, MobileNet had the best performance in training accuracy and validation accuracy (99.30% and 98.89%, respectively).

Further, [7] analysed and identified four different algorithms for adapting the OCR process to recognise Tamil scripts, and support vector machines (SVM) was identified as the best option. Meanwhile [8] a method called advanced maximally stable extremal regions (AMSER) to improve the accuracy of identifying Tamil characters in images by detecting the extremal region and characteristic function, the accuracy of recognition of which was 95.59%. This method was introduced because of the low accuracy of inscriptions images using OCR. In addition, [9] suggested a k number of clusters (k-means clustering) for an ancient Kannada text using scale-invariant Fourier transform (SIFT) and speeded up robust features (SURF). Moreover, in [10], OCR was used to identify ancient Tamil inscriptions on stone using a feature extracted with the SIFT algorithm.

According to [11], who employed the Siamese network in few-shot learning (FSL),the local feature extraction of Chinese characters performs better using the VGG16 network as the backbone feature extraction network, achieving a recognition accuracy of 82.67%. Meanwhile, [12] the efficient and accurate scene text (EAST) detection model and the feature extractor VGG16 to extract painting inscriptions by combining the characteristics of Chinese paintings, achieving a high accuracy of 89%. In addition, in [13], Sundanese writing inscribed on palm leaves was recognised using a three-layer CNN with a 73% recognition accuracy. Next, in [14], was employed for recognition, and Tesseract training was performed using a deep neural network architect. Then, CNN long–short-term memory (LSTM) networks were configured and trained for the language model of the Tamizhi script, leading to an OCR accuracy of 91.21%. Furthermore, [15] used a CNN to extract and translate information from each character into Modern Tamil, achieving an accuracy of 94.6%.

CNNs are used as feature extractors, as well as classifiers, for their ability to recognise 33 classes of basic characters from Devanagari ancient manuscripts, as in [16], reaching a recognition accuracy of 93.73%. Further, the historical Kannada handwritten characters are recognised using the line segmentation approach with LBP features in [17], and the SVM classifier achieved a good performance, with an accuracy of 96.4%. In [18], a CNN was used with dropout to recognise Brahmi words, with a 92.47% accuracy. As well, [19] designed and developed an automatic recognition tool for variant characters to assess tablet inscriptions, leading to an accuracy of the trained model ResNet50-18 of 90%. In addition, [20] utilised CNN and MobileNet to detect Tamil-Brahmi script, achieving an accuracy of 68.3%, but MobileNet outperformed all other models employed. In [21], a CNN was used to classify and recognise Brahmi characters, as well as broken part of these characters, and VGG16 models provided results with the best accuracy, at 93.33%.

Through this literature review, various approaches, such as SIFT, ResNet18 and VGG16, for feature extraction and classification techniques, including CNN, were explored. The results reported in these papers vary depending on such factors as the condition of the inscriptions and the techniques employed. Despite an extensive review, it is noteworthy that no research paper was identified that specifically addresses the recognition and translation of inscriptions in the ancient Southern Arabic Musnad font.

III. Dataset

A. Dataset Collection and Description

A request was made to the Ministry of Culture to provide a collection of images for use as a dataset in this work, and they did so with a collection of images from the Heritage Commission of the Ministry of Culture. The dataset contained images of the Musnad inscriptions in Lahyani form. And contained images of the Musnad from Qaryat al-Faw, Yemen and Al-Ula, totalling 293 inscriptions, which were divided into three categories: real, written and anointed, as shown in Fig. 3.

As the Musnad inscriptions span consecutive historical stages, researchers divided them into three types: ones that take on a geometric look, as in Fig. 4(b); ones that tend to bend, as in Fig. 4(c); and ones that appear exaggerated in decoration, as in Fig. 4(a) [2][22]. Further, the dataset comprises all feasible images sourced from the Ministry of Culture, ensuring the highest quality.

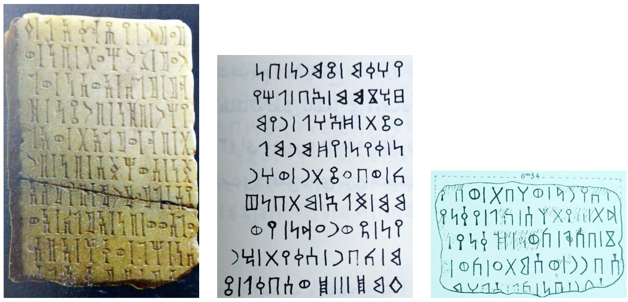

(a) Real (b) Written (c) Anointed

Fig. 3. The three categories of the dataset.

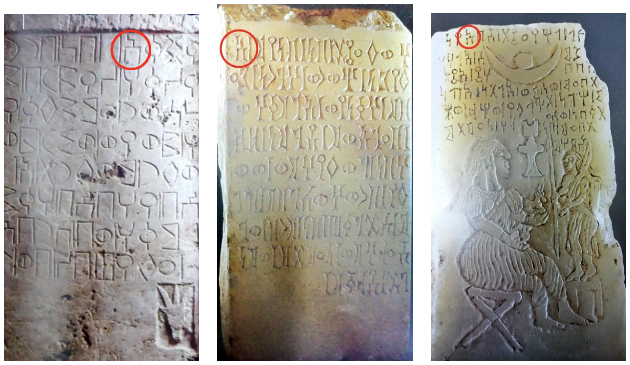
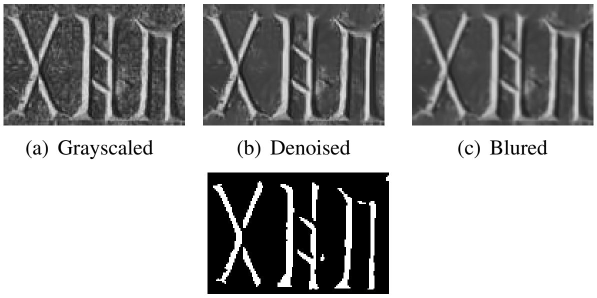

(a) Decorated (b) Geometric (c) Bend

Fig. 4. Three periods of the musnad inscriptions.

B. Image Preprocessing

The preprocessing experiments aimed to improve the recognition efficiency and accuracy of the dataset [15]. Of the several experiments conducted, only two produced good results, one for natural scene images and one for anointed and written images, as described below:

1) Preprocessing for natural scene images: In the first experiment, four processes were carried out using natural scene images. After first converting the input image from RGB to grayscale, as shown in Fig. 5(a), the grayscale image was then denoised using the cv2.fastNlMeansDenoising function, and the result of the denoising is shown in Fig. 5(b). The denoised grayscale image is then smoothed with the cv2.GaussianBlur function, as shown in Fig. 5(c). Finally, the smoothed grayscale image is then binarised using the cv2.threshold function with a threshold of 130, as shown in Fig. 5(d).

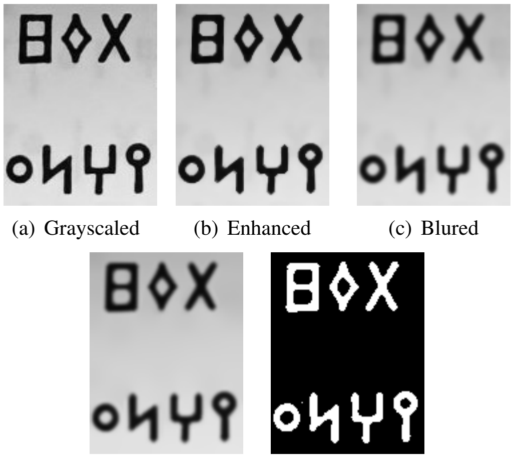

(a) Grayscaled (b) Denoised (c) Blured

(d) Binarized

Fig. 5. Preprocessing for natural scene images.

2) Preprocessing for anointed and written images: In the second experiment, five processes were carried out using the anointed and written images, starting by converting the input image from RGB to grayscale, as shown in Fig. 6(a).

Then, the cv2.medianBlur function was used to remove noise from the grayscale image, as shown in Fig. 6(b), after which the median filter result was smoothed using the cv2.GaussianBlur function with a kernel size of 5, as shown in Fig. 6(c). Thereafter, the Gaussian blur result was further denoised using the cv2.fastNlMeansDenoising function, as shown in Fig. 6(d). Finally, the denoised result was binarised using the cv2.adaptiveThreshold function with a Gaussian method and the inverse binary thresholding method with a threshold value of 255, as shown in Fig. 6(e).


(a) Grayscaled (b) Enhanced (c) Blured

(d) Denoised (e) Binarized

Fig. 6. Preprocessing for anointed and written images.

C. Segmentation and Augmentation

Roboflow Datasets is a comprehensive computer vision tool that offers a range of functionalities, including dataset upload, organization, collaboration, labeling, augmentation, and processing [23]. The selection of the Roboflow Datasets tool is based on its notable ease of use, particularly in scenarios where characters appear too close to each other or when certain images exhibit characters that have been damaged. In Fig. 7, the labeling process is illustrated, followed by the implementation of a data augmentation technique involving a rotation of 10 degrees in both clockwise and counterclockwise directions. This technique was chosen for its recognized capacity to diversify the dataset, introducing variations in orientation that enhance recognition accuracy.

Subsequently, the dataset was divided into two sections, with 80% allocated for training and 20% for testing. The training set consisted of 11,103 images, while the test set comprised 2,763 images. Fig. 8 illustrates the distribution of the dataset across 29 classes. To ensure consistency, all images in the dataset were standardized to a size of 224 x 224 pixels.

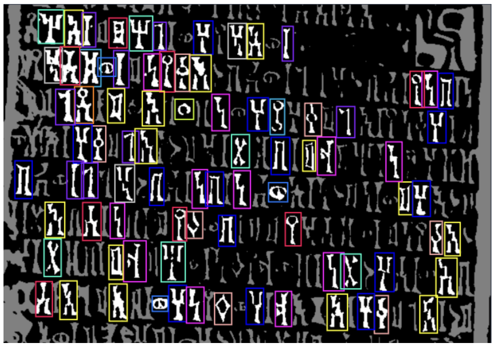


Fig. 7. Segmentation result.

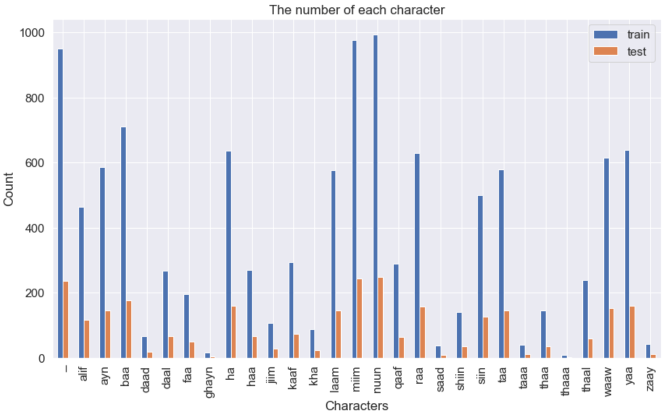


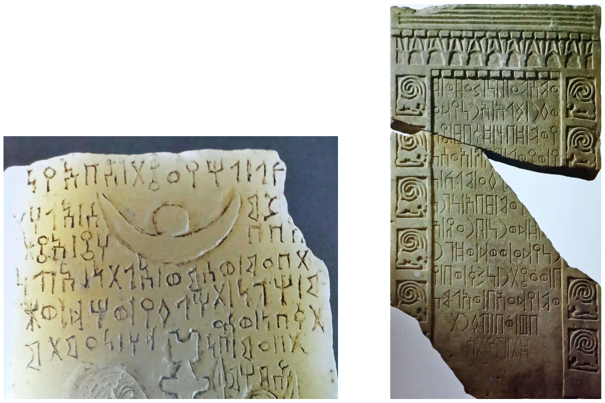
Fig. 8. Musnad characters classes.

D. Challenges and Limitations

Recognising the Musnad font involves challenges for several reasons. First, the segmentation process requires manual intervention, as it becomes difficult due to complex structures, as several images contain illustrations, as shown in Fig. 9(a). Further, some characters were too close to each other, some were not on the same line and some were cut off due to broken backgrounds, as shown in Fig. 9(b). Furthermore, Fig. 10(b) displays the results of preprocessing, where certain characters lack clarity due to varied surface conditions, as depicted in Fig. 10(a).

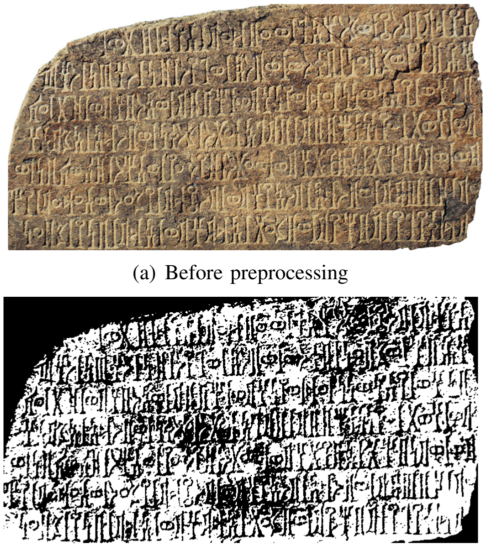
IV. Methodology

Here, the key steps and techniques implemented in the work to accomplish the desired outcomes. First, an image is captured and uploaded, after which the system implements image preprocessing. Third, the system detects and recognises the correct Musnad characters, and finally, the Musnad characters are translated into Arabic characters. For reference, Fig. 11 shows the proposed workflow of the system. In subsequent sections, each of these processes will be discussed in detail, providing a comprehensive understanding of the methodology employed.



(a) Inscription containing illustrations (b) Inscription with a broken background

Fig. 9. Dataset sample.



(a) Before preprocessing

(b) After preprocessing

Fig. 10. Surface condition: before and after preprocessing.

A. Image Preprocessing

The process for preprocessing natural scene images mentioned in III-B1 will be applied to the input image, with the addition of a new dilation process by utilising the cv2.dilated function.

B. Text Detection and Recognition

1) Contour Detection: Contour detection is a method often used in computer vision and image processing to detect and

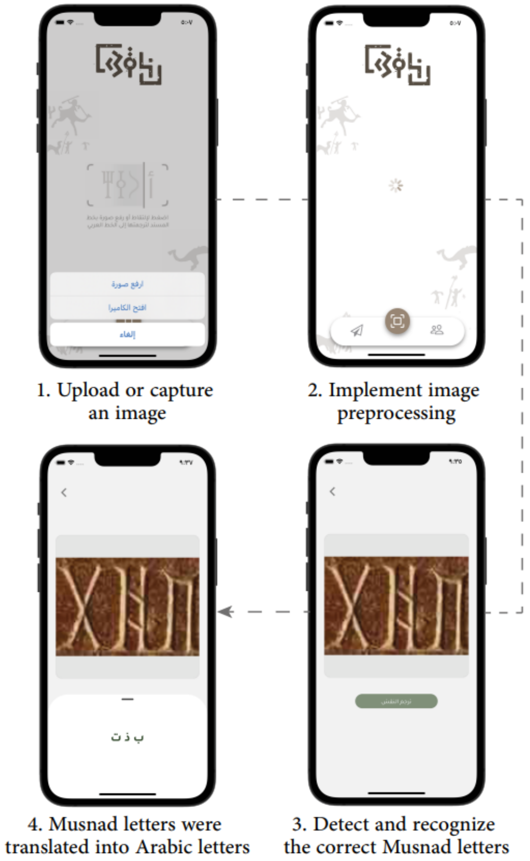

ارفع صورة
افتح الكاميرا
إلغاء
1. Upload or capture an image
2. Implement image preprocessing
ب ذ ت
4. Musnad letters were translated into Arabic letters
3. Detect and recognize the correct Musnad letters

Fig. 11. Workflow of the system.

find the outlines of objects in an image [24]. Unfortunately, the project's pretrained recognition models were only capable of correctly identifying one character, as they treated the image as containing only one character, presneting an obstacle in the recognition process. This project used a method for dealing with this problem by utilising the OpenCV function, which provides two functions: findContours and drawContours. The contouring process utilised herein is findContours, which detects the contours of the inscription to recognise all the characters of the Musnad inscriptions [25].

The findContours function requires three arguments: IMAGE, RETR_EXTERNAL and CHAIN_APPROX_SIMPLE, all of which were carefully chosen to optimise the contour detection process. By applying this method, the Musnad inscriptions in the image were successfully outlined, as demonstrated in Fig. 12.

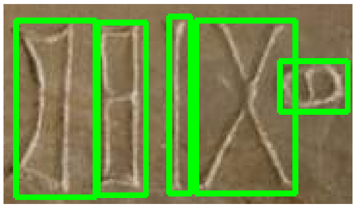


Fig. 12. Contour detection example.

2) *Deep Learning Models:* CNN is a highly efficient technique for image classification, and it is widely used in many recognition problems. CNN can perform both feature extraction and classification [26]. However, deep learning algorithms typically require more time and data than conventional machine learning systems offer to achieve an optimal performance [27]. As such, transfer learning is a technique that leverages a model pretrained on a specific dataset and that adjusts its parameters to suit new datasets. This approach is more efficient than creating a new CNN model from scratch [27], but one challenge that can arise is overfitting. In this case, the model becomes too specialised and cannot be adapted to new data [27]. To overcome this issue in this paper, early stopping and dropout techniques have been employed. Thus, this section presents how to employ the concept of transfer learning using feature extraction with VGG16, ResNet-50 and MoileNetV2, which were chosen due to their proven effectiveness in identifying ancient inscriptions [21], [6], [19].

VGG16: VGG16 is a CNN model for image classification developed by the Visual Geometry Group (VGG), comprising 16 layers in total, including 13 convolutional layers and three fully connected layers, using only 3×3 convolutional layers stacked atop each other [28]. The VGG16 model in this project is composed of two parts: a pretrained VGG16 model and additional custom layers added atop the pretrained model. Further, the VGG16 model is loaded as a pretrained model with the input shape set to (224,224,3), the top layer set to 'false' and the weights set to 'none'. By setting the top layer to false, the VGG16 model's fully connected layers are not included, allowing the model to be used as a feature extractor. After loading the pretrained model, a new sequential model is created, and the pretrained model is added as the first layer. The output of the pretrained VGG16 model is flattened using a flattened layer, after which the model has two fully connected Dense layers with 256 and 512 neurons, respectively. Batch normalisation is then applied after each Dense layer, followed by the rectified linear unit (ReLU) activation function. Dropout with a rate of 0.25 was applied after each activation function, producing an output Dense layer with 29 classes and a softmax activation function.

ResNet50: ResNet50: ResNet-50 is a CNN model with 50 convolutional layers that contain residual blocks, as well as approximately 25.6 million parameters [29]. Both the ResNet-50 and VGG16 models underwent the same modification to their fully connected layer (FCL) architecture, utilising the same parameter values.

MobileNetv2: MobileNetV2 is a CNN model designed to be fast and efficient to reduce the large network size and to minimise the cost of network computing [30]. The MobileNetV2 has undergone the same modifications as the VGG16 and Resnet-50 models, but they resulted in an underwhelming performance. Therefore, to improve the results, we altered the FCL modifications to differ from those applied to the VGG16 and ResNet-50 models to enhance its effectiveness.

The MobileNetV2 model is comprised of two parts: a pre-trained MobileNetV2 model and additional custom layers added atop the pre-trained model. In addition, the pre-trained MobileNetV2 model is loaded with ImageNet weights, and the input shape is set to (224,224,3) with the top layer set to ‘false’. A new sequential model is created, with the pretrained MobileNetV2 model being added as the first layer. Then, the output from the MobileNetV2 model is processed through a GlobalAveragePooling2D layer, after which it is flattened using a Flatten layer. It is then fed into three Dense FCLs with 256 neurons and a ReLU activation function, followed by a Dropout layer with a rate of 0.25. Finally, the output layer is a Dense layer with 29 classes and a softmax activation function.

Experimental Setup and Conditions: The experiments were performed on a computer system equipped with the Windows 11 Pro operating system, an Intel Core i5 11400 central processing unit (CPU) and an Intel(R) UHD Graphics processing unit (GPU). In addition, the machine learning framework used was Keras 2.11.0, which was run using Python 3.9.12 in Spyder Anaconda.

C. Evaluation Metrics

Evaluation metrics are used to measure the performance of a classifier with a dataset, providing a way of measuring the model’s performance in making correct predictions. The confusion matrix provides information about a predictive model’s performance, and various metrics can be derived therefrom to gain deeper insights. These metrics include accuracy, precision, recall, F1 score and the confusion matrix itself [31]. Below are the evaluation metrics used to evaluate the models herein:

Confusion matrix: Compares predictions made by the trained classifier to the true labels in the test set. Correctly detected results are shown in diagonal cells, while incorrect predictions are shown in the remaining cells [32].

Accuracy: This metric establishes the model’s accuracy in making predictions, as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Precision: Indicates the proportion of correctly positive predictions of all positive predictions. This metric establishes the reliability of the predictive model in making accurate predictions, as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall (also known as Sensitivity or True Positive Rate): Applies the same principle as precision, but instead of focusing on false positives, it focuses on false negatives.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1-Score: This metric assesses the accuracy with which the predictive model can predict positive values while accounting for both FN and FP. It provides a balance between the two measures by taking the harmonic mean of precision and recall [31].

$$F = \frac{2TP}{2TP + FP + FN} \tag{4}$$

V. Result and Discussion

We evaluated the performance of pretrained models, namely VGG16, ResNet50 and MobileNetV2, in recognising Musnad inscription characters. The VGG16 model underwent training for 63 epochs with early stopping enabled, which took approximately 12.42 hours to complete. Similarly, the ResNet50 model was trained for 78 epochs, which required approximately 8.08 hours of training time. Finally, the MobileNetV2 model underwent training for 149 epochs, which required approximately 5.38 hours to train. For all models, the learning parameters remained consistent, including a learning rate of 0.00001, 150 epochs with early stopping and a batch size of 32.

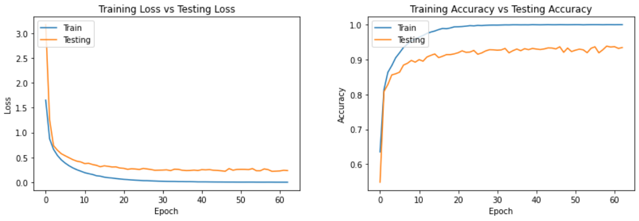
The model evaluations were based on various metrics, as mentioned in IV-C, and the results are summarised in Table I. The VGG16 model demonstrated the best performance among the tested models, achieving an average accuracy recognition rate of 93.81% across 29 character classes. Both the ResNet50 and MobileNetV2 models also showed excellent accuracy results, although they were slightly lesser compared to the VGG16 model.

Fig. 13 demonstrates the convergence and effectiveness of the VGG16 model by illustrating its loss and accuracy during the training and testing phases, providing valuable insights into the model’s performance. Likewise, Fig. 14 and Fig. 15 display the loss and accuracy trends of the ResNet50 and MobileNetV2 models, respectively, highlighting their training progress and overall performance and contributing a comprehensive view of how these models perform over time. Moreover, Fig. 16 presents the confusion matrix for all three models, providing an insightful visualisation of their classification performance.

These findings suggest VGG16 can effectively recognise Musnad inscription characters with high accuracy. As such, the results demonstrate the potential of this model in the field of character recognition and provide valuable insights for further improvements and applications.

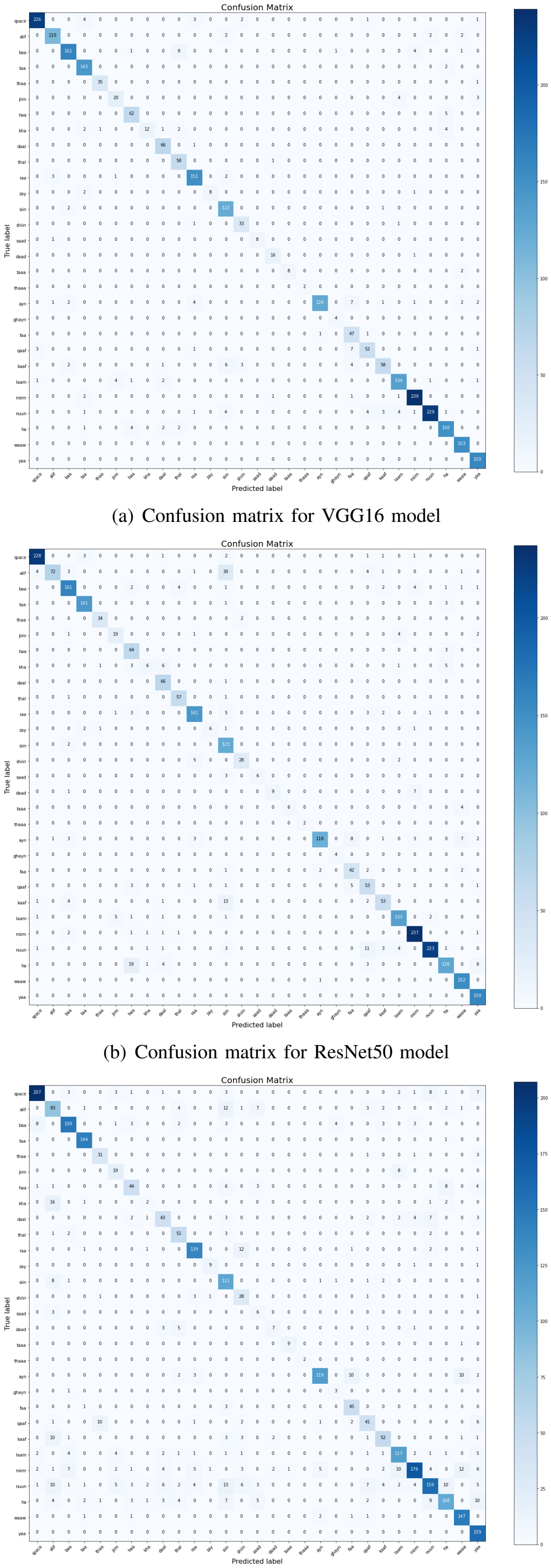
TABLE I. Transfer Learning Results

	Precision	Recall	F1-score	Accuracy
VGG16	93.00	91.00	91.00	93.81
ResNet50	90.00	83.00	85.00	89.39
MobileNetV2	73.00	77.00	74.00	80.02



(a) Training Loss vs Testing Loss (b) Training Accuracy vs Testing Accuracy

Fig. 13. Loss and accuracy plot for VGG16.

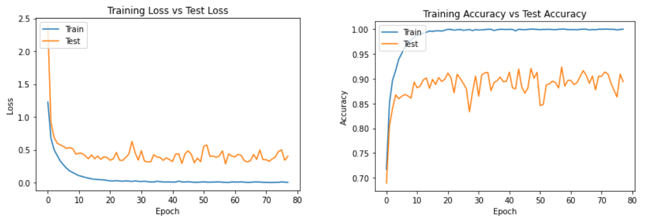


(a) Confusion matrix for VGG16 model

(b) Confusion matrix for ResNet50 model

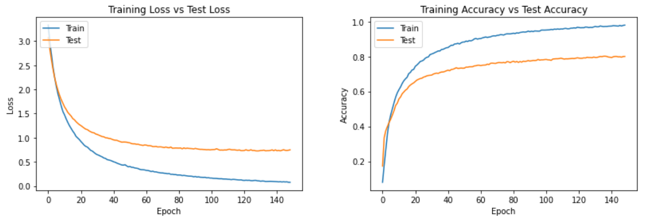
(c) Confusion matrix for MobileNetV2 models

Fig. 16. Confusion matrix for all three models.



(a) Training Loss vs Testing Loss (b) Training Accuracy vs Testing Accuracy

Fig. 14. Loss and accuracy plot for ResNet50.



(a) Training Loss vs Testing Loss (b) Training Accuracy vs Testing Accuracy

Fig. 15. Loss and accuracy plot for MobileNetV2.

VI. Conclusion

This paper automated the recognition and translation of Musnad inscriptions, making a unique contribution to ancient text recognition and translation by creating a Musnad inscription dataset, developing a recognition and translation system and comparing three CNN models (VGG16, MobileNetv2 and ResNet-50). The paper's findings indicate that real-scene images of the language's inscriptions can form a useful dataset, that the optimal sequence of image processing techniques to improve recognition accuracy and that the VGG16 deep learning model provides the highest accuracy, with a recognition rate of 93.81%. Further, ResNet-50 achieved a recognition rate of 89.39% and MobileNetV2 a rate of 80.02%. In the future, the objective will be to enhance the recognition method for practical use in real-world scenarios. As such, the focus will eventually shift towards addressing broken inscriptions and recognising old drawings, ensuring the effectiveness of techniques in challenging scenarios and with unseen or difficult-to-read inscriptions, including those with varying levels of degradation, different lighting conditions or unconventional surfaces. This enhancement will be achieved by enlarging the datasets for model training, improving image processing to generate a high-quality dataset and using NLP to determine the Arabic meanings of the words in the Musnad inscriptions.

Acknowledgment

We would like to thank the Saudi Ministry of Culture for providing us with the Musnad inscription images, which have been essential to creating our dataset. As well, we deeply appreciate the Ministry's support in preserving and sharing these valuable resources.

REFERENCES

[1] A. Saleh, "History of the arabian peninsula," 2010.

[2] S. Alyaarubi, "A brief study on musnad font," 2013.

[3] I. bin Nasser Al-Braihi, "Crafts and industries in the light of south musnad inscriptions," 2000.

[4] A. bin Ali Abu Hadra, "In the language of the yemeni people," 2013.

[5] R. Santoso, Y. K. Suprapto, and E. M. Yuniarno, "Kawi character recognition on copper inscription using yolo object detection," in *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*. IEEE, 2020, pp. 343–348.

[6] Y. Fujikawa, H. Li, X. Yue, C. Aravinda, G. A. Prabhu, and L. Meng, "Recognition of oracle bone inscriptions by using two deep learning models," *International Journal of Digital Humanities*, pp. 1–15, 2022.

[7] M. Rajkumar, R. A. B., and v. Janhavi, "Analyzing different algorithms and techniques to find optical character recognition for tamil scripts," 2020.

[8] A. N. Kumar and G. Geetha, "Character recognition of ancient south indian language with conversion of modern language and translation," *Caribb. J. Sci*, vol. 53, no. 20, pp. 2019–2031, 2019.

[9] P. Ravi, C. Naveena, and Y. Sharathkumar, "Ocr for historical kannada documents using clustering methods," *Indian Journal of Science and Technology*, vol. 13, no. 35, pp. 3652–3663, 2020.

[10] M. Merline Magrina and M. Santhi, "Ensemble classifier system for offline ancient tamil character recognition," *SSRG International Journal of Electronics and Communication Engineering (SSRG-IJECE)*, 2019.

[11] M. Wang, Y. Cai, L. Gao, R. Feng, Q. Jiao, X. Ma, and Y. Jia, "Study on the evolution of chinese characters based on few-shot learning: From oracle bone inscriptions to regular script," *PloS one*, vol. 17, no. 8, p. e0272974, 2022.

[12] S. Zhai, L. Liao, Y. Lin, and L. Lin, "Inscription detection and style identification in chinese painting," in *2020 Chinese Automation Congress (CAC)*. IEEE, 2020, pp. 7434–7438.

[13] S. Chadha, S. Mittal, and V. Singhal, "Ancient text character recognition using deep learning," *International Journal of Engineering Research and Technology*, vol. 3, no. 9, pp. 2177–2184, 2020.

[14] M. Munivel and V. Enigo, "Optical character recognition for printed tamizhi documents using deep neural networks." *DESIDOC Journal of Library & Information Technology*, vol. 42, no. 4, 2022.

[15] S. Subadivya, J. Vigneswari, M. Yaminie, and M. Diviya, "Tamilbrahmi script character recognition system using deep learning technique," *International Journal of Computer Science and Mobile Computing*, vol. 9, no. 6, pp. 114–119, 2020.

[16] S. R. Narang, M. Kumar, and M. K. Jindal, "Deepnetdevanagari: a deep learning model for devanagari ancient character recognition," *Multimedia Tools and Applications*, vol. 80, no. 13, pp. 20 671–20 686, 2021.

[17] P. Bannigidad and C. Gudada, "Historical kannada handwritten scripts recognition system using line segmentation with lbp features."

[18] N. Gautam, S. S. Chai, and J. Jose, "Recognition of brahmi words by using deep convolutional neural network," 2020.

[19] S. Luo, "Tablet inscription recognition based on the convolutional neural network," *Journal of Chinese Writing Systems*, vol. 6, no. 3, pp. 185–197, 2022.

[20] S. Dhivya and U. G. Devi, "Tamizhi: Historical tamil-brahmi script recognition using cnn and mobilenet," *ACM TRANSACTIONS ON ASIAN AND LOW-RESOURCE LANGUAGE INFORMATION PROCESSING*, vol. 20, no. 3, 2021.

[21] K. N. Wijerathna, R. Sepalitha, T. Indika, H. Athauda, P. Suranjini, J. Silva, and A. Jayakodi, "Recognition and translation of ancient brahmi letters using deep learning and nlp," in *2019 International Conference on Advancements in Computing (ICAC)*. IEEE, 2019, pp. 226–231.

[22] A. Makiash, "Southern arabic script: Its origins, spread, and relationship with northwestern arabic script," vol. 12, no. 12, pp. 338–366, 2009.

[23] Roboflow, "Roboflow docs," https://docs.roboflow.com/, 2023.

[24] H. Sidhwa, S. Kulshrestha, S. Malhotra, and S. Virmani, "Text extraction from bills and invoices," in *2018 international conference on advances in computing, communication control and networking (ICACCCN)*. IEEE, 2018, pp. 564–568.

[25] P. Manuaba and K. A. T. Indah, "The object detection system of balinese script on traditional balinese manuscript with findcontours method," *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 11, no. 3, pp. 177–184, 2021.

[26] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[27] J. Pardede, B. Sitohang, S. Akbar, and M. L. Khodra, "Implementation of transfer learning using vgg16 on fruit ripeness detection," *Int. J. Intell. Syst. Appl*, vol. 13, no. 2, pp. 52–61, 2021.

[28] S. Tammina, "Transfer learning using vgg-16 with deep convolutional neural network for classifying images," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 10, pp. 143–150, 2019.

[29] V. S. Dhaka, S. V. Meena, G. Rani, D. Sinwar, M. F. Ijaz, and M. Woźniak, "A survey of deep convolutional neural networks applied for prediction of plant leaf diseases," *Sensors*, vol. 21, no. 14, p. 4749, 2021.

[30] S. METLEK, "Disease detection from cassava leaf images with deep learning methods in web environment," *International Journal of 3D Printing Technologies and Digital Industry*, vol. 5, no. 3, pp. 625–644, 2021.

[31] S. Kok, A. Azween, and N. Jhanjhi, "Evaluation metric for crypto-ransomware detection using machine learning," *Journal of Information Security and Applications*, vol. 55, p. 102646, 2020.

[32] Z. Omiotek and A. Kotyra, "Flame image processing and classification using a pre-trained vgg16 model in combustion diagnosis," *Sensors*, vol. 21, no. 2, p. 500, 2021.

Learnable Local Similarity for Face Forgery Detection and Localization

Lingyun Leng[1], Jianwei Fei[2], Yunshu Dai[3]
College of Cyber Security, Jinan University, Guangzhou, 510632, Guangdong, China[1]
School of Computer, Nanjing University of Information Science & Technology, Nanjing, 210044, Jiangsu, China[2]
School of Cyber Science Technology, Sun Yat-sen University, Shenzhen, 518107, Guangdong, China[3]

***Abstract*—The emergence of many face forgery technologies has led to the widespread of forgery faces on the Internet, causing a series of serious social impacts, thus face forgery detection technology has attracted increasing attention. While many face forgery detection algorithms have demonstrated impressive performance against known manipulation methods, their efficacy tends to diminish severely when applied to unknown forgeries. Previous research commonly viewed face forgery detection as a binary classification problem, disregarding the crucial distinction between real and forged faces, thereby limiting the generalizability of detection algorithms. To overcome this issue, this paper proposes a novel face forgery detection method that utilizes a trainable metric to learn local similarity between local features of facial images, achieving a more generalized detection result. What's more, it incorporate cross-level features to accurately locate forgery regions. After conducting extensive experiments on FaceForensics++, Celeb-DF-v2, and DFD, which demonstrate that the effectiveness of the proposed method is comparable to state-of-the-art detection algorithms.**

Keywords*—*Face forgery detection; local similarity; forgery localization; generalized detection

I. Introduction

With the advancements in computer vision and deep learning technology, face forgery technologies have become a growing concern for society. These technologies have matured significantly in recent years, producing counterfeit faces that are indistinguishable to the human eye [1], [2], [3], [4]. Criminal misuse of these technologies can pose severe societal consequences, including pornographic featuring public figures, the spread of political misinformation, and fraudulence that jeopardizes personal and property rights. In response to this challenge, many researchers have developed studies on face forgery detection approaches [5], [6], [7]. Though these approaches perform well in specific scenarios, they are often inadequate when it comes to detecting forgeries unseen in the training data, known as the generalization problem.

Early research considered face forgery detection a binary classification task, employing convolutional neural networks (CNN) to distinguish real and forged faces. While such approaches have shown impressive performances in in-domain settings where training data and testing data are forged by the same algorithm, they cannot be easily applied to unknown domains given that the unknown forgery algorithms have different forgery features. Consequently, some face forgery detection approaches have been proposed to mine generalizable artifact traces [8], [9], [10]. These approaches rely on identifying discrepancies in image features such as brightness, color, and texture to distinguish forged faces. However, they can be affected by the quality of the images and thus may not be suitable for real-world scenarios. To address this issue, emerging approaches centered around data augmentation are being employed [11], [12]. By using various forgery algorithms, these techniques aim to expand the training data, to improve generalization capability. However, it's worth noting that training data depending on forgery detection approaches may become ineffective in light of the continuous advancements in forgery algorithms. On the other hand, some researchers concentrate on the identification details of forged faces, discover identity discrepancies to identify forged faces, and have achieved remarkable success with face replacement [13], [14]. Nevertheless, this type of method cannot detect face images with unchanged identity information. As face forgery algorithms continue to advance, the forgery traces have become increasingly subtle and difficult to detect using previous approaches. Thus, some researchers start paying attention to identifying fine-grained local feature inconsistencies [15], [16], [17].

In this paper, we focus on face forgery detection approach that does not rely on data augmentation but extracts essential features of real and forged faces. Our solution is based on the observation that real faces typically exhibit evenly distributed features and local region similarities [15], whereas forged faces usually exhibit local abnormalities resulting from the blending between real and forged regions. Inspired by this, we propose a novel generalized forgery face detection approach. Our approach improves generalization by leveraging auxiliary constraints between local features of real and forged faces. Unlike existing methods that rely on metrics such as cosine similarity [17], we utilize a learnable network to measure the similarities between local features, which makes the similarity measurement better suited to aligning features extracted by CNN backbones. We also introduce a cross-level forgery localization module that integrates various features across different levels with a lightweight attention module. Our approach enables accurate forgery localization and forged face recognition with strong generalizability. In brief, our contributions are summarized as follows:

- We propose to learn the dense fine-grained similarity between real and forged local features, which greatly improves the generalization of face forgery detection approach.
- We propose a Y-shaped network that achieved accurate face forgery detection and localization with the proposed cross-level attentional feature fusion module.

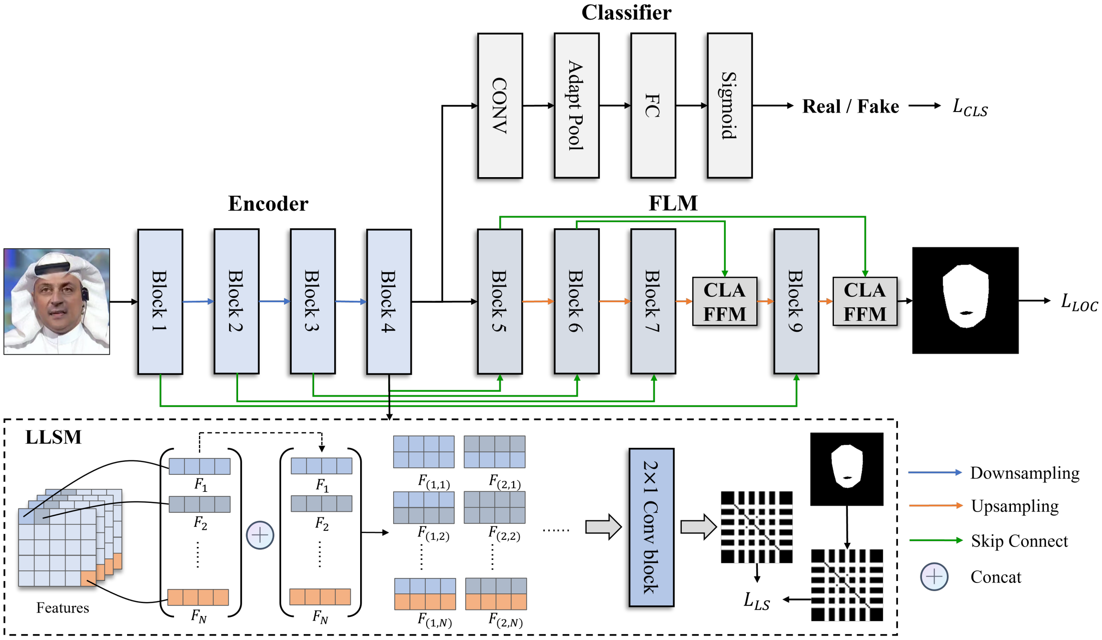


Fig. 1. The overall architecture of the proposed approach.

- Extensive experiments validate the superior detection performance and impressive generalization ability of our approach.

II. Related Work

A. Face Forgery Algorithms

Face forgery technology has advanced quickly in recent years. It can be categorized into three categories: face swap, face editing, and face generation, based on the forgery objects. Early face swap algorithms are mostly accomplished using graphics techniques [18], which are complex and challenging. With the rapid development of deep learning, several novel face swap algorithms have emerged, significantly reducing the difficulty of face swapping [3], [4]. The advance of Generative Adversarial Networks (GAN) has further improved the realism of the forged faces [1], [2].

B. Face Forgery Detection Algorithms

Early face forgery algorithms were not very mature, resulting in flawed faces with obvious artifacts. Therefore, early detection algorithms depend mostly on CNN to catch these probable artificial traces, such as color artifacts [8], blink rate [19], head position [14], synthetic artifacts [9], and so on. However, with the advance of face forgery algorithms, these problems have largely been resolved, forcing researchers to develop new detection methods capable of identifying generic forgery clues.

On the one hand, the researchers discovered that forgery faces have invariable forgery patterns in the frequency domain, giving rise to frequency-aware forgery detection approaches. Qian *et al.* [20] proposed a two-stream collaborative learning framework that leverages frequency-aware image decomposition and local frequency statistics to extract forgery patterns. Meanwhile, Chen *et al.* [21] tackled face forgery detection through local relational learning, integrating RGB and frequency information using an attention module to improve generalization. On the other hand, temporal inconsistencies of forged videos have become a significant clue for forgery detection. Time-aware models that extract temporal features from numerous single-frame inputs have been introduced to detect the authenticity of face videos. Güera *et al.* [22] used Long Short-Term Memory (LSTM) to extract temporal features from numerous single-frame. Gu *et al.* [23] detected local dynamic inconsistencies induced by tiny movements in forged videos. In addition, approaches based on advanced semantic information are proposed. Haliassos *et al.* [24] suggested detecting forgery videos by analyzing differences in lip movements between real and forgery videos, while Dong *et al.* [13] developed an identity consistency transformer to safeguard celebrities by detecting identity inconsistencies between inner and outer faces. However, approaches based on common forgery clues may not be suitable for real-world scenarios with unknown forgery clues, and approaches based on temporal inconsistency may be limited by video quality.

Due to the flaws in splicing, blending, and editing procedures present in the majority of available face forgery algorithms, forged faces may contain features from multiple sources, leading to local inconsistencies. Shang *et al.* [15] proposed to capture pixel-level and region-level differences for face forgery detection. Zhao *et al.* [16] proposed pairwise self-consistency learning of local features, which achieved excellent performance in generalization. Sun *et al.* [25] enhanced the generalization by creating positive and negative sample pairings and contrastively learning real and forged regions. The studies mentioned above highlight the importance of local inconsistency in improving the generalization of face forgery detection. They mainly use fixed metric measures when calcu-

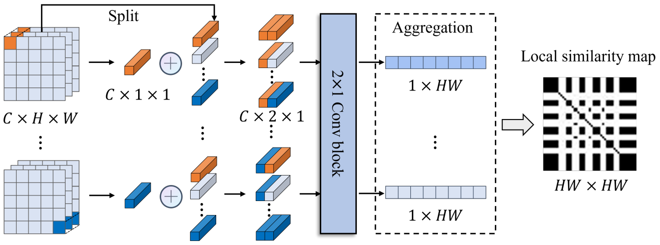


Fig. 2. The proposed learnable local similarity module (LLSM).

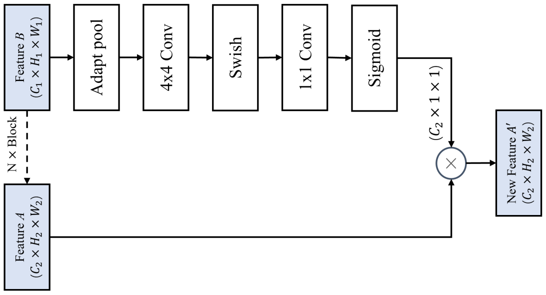

Feature B $(C_1 \times H_1 \times W_1)$
Adapt pool
4x4 Conv
Swish
1x1 Conv
Sigmoid
$(C_2 \times 1 \times 1)$
N × Block
Feature A $(C_2 \times H_2 \times W_2)$
New Feature A' $(C_2 \times H_2 \times W_2)$

Fig. 3. The proposed cross-level attentional feature fusion module (CLAFFM).

lating consistency or similarity, which may result in feature misalignment issues. For example, when using normalized cosine similarity to measure the similarity of local features, it directly transforms the feature onto a unit hypersphere. However, the most generalizable features extracted from the backbone network may not satisfy the similarity requirement measured by the cosine similarity metric.

In this paper, we propose a plug-and-play learnable local similarity module that densely imposes fine-grained similarity constraints on the features extracted from the backbone network. Unlike existing approaches, the similarities are not directly *calculated* but learned using ConvNets, and we turn the optimization target into binary classification between features.

III. Methodology

A. Overview

The proposed approach is shown in Fig. 1. It is a Y-shape network with an encoder, a decoder, and a classifier head, where a learnable local similarity module works in collaboration with the encoder, a cross-level attention feature fusion model connects the encoder and decoder that performs pixel-level forgery localization. The classifier head implements image-level binary classification of authenticity (real/fake) for the input faces.

B. Learnable Local Similarity Module

Deepfake faces contain subtle artifacts or feature conflicts from several sources, this difference is frequently displayed at the fine-grained feature level and is difficult to detect. Therefore, we build a specific fine-grained local feature similarity map to mine fine-grained feature inconsistencies by computing its feature similarity with all locations based on local features. We constructed a learnable local similarity module for the augmented model to further enhance the model's capacity for generalization by mining fine-grained local differences.

The learnable local similarity module (LLSM) is a plug-and-play module that takes as input the deep features extracted by the backbone and calculates the similarity between local features. Given a face image I, we first pass through the encoder to obtain its middle layer feature $F \in R^{H \times W \times C}$ where C, H, and W represent channel, length, and width respectively. The middle layer feature F is then fed into LLSM, which predicts a single-channel local similarity map. As shown in Fig. 2, the 3-d tensor F is first unfold into a 2-d matrix of shape $\mathbb{R}^{(H \times W) \times C}$, $F_{(i,j)}$ $(0 \leq i < H, 0 \leq j < W)$ denote the local feature, where i and j are the spatial index. A shallow ConvNet with a single kernel of size $2 \times 1 \times C$, denoted by f then perform convolution operation on each local feature pairs. For $\forall i, m \in [1, H], \forall j, n \in [1, W]$, to predict the local similarity map, we use the following equation to calculate the similarity of any local features on the feature F:

$$M_{i*H+j,m*H+n} = \sigma(f(F_{(i,j)}, F_{(m,n)})), \tag{1}$$

where M is the predicted 2-d map of size $\mathbb{R}^{(H \times W) \times (H \times W)}$. That means each local feature is compared with all local features including itself by f, which outputs a binary prediction on the similarity of arbitrary two local features. After all the local features are calculated and arranged according to the corresponding positions, we can obtain a feature inconsistency map M^F with the shape of $(HW) \times (HW)$. On the optimization strategy for the feature inconsistency map, we calculating the ground truth local similarity map are as follows: First, downsample the mask until its dimensions match those of the middle layer feature map. Then expand the mask map into a one-dimensional tensor $m^{'}$, calculate the Cartesian product of all positional features, and return the local similarity map of $(HW) \times (HW) \times 2$ after adjusting the shape. After that, it is divided into two identically sized feature maps m_1,m_2 based on the third dimension, and the ultimate ground truth local similarity map M^{GT} is obtained by binarizing the two after an absolute value difference. We adopt the following BCE loss to supervise the training:

$$M^F = LSC\left(F\right), \tag{2}$$

$$m_1, m_2 = split\left(Cartesia\_prod\left(m^{'}, m^{'}\right)\right), \tag{3}$$

$$M^{GT} = binary\left(\left|m_1 - m_2\right|\right), \tag{4}$$

$$L_{LS} = \frac{1}{N}\sum_{i=1}^{N} BCE\left(M_i^F, M_i^{GT}\right), \tag{5}$$

where LSC stands for the method of predicting the local similarity map, *split* means split features by dimension, *Cartesia_prod* means Cartesian product operation, and *binary* means binarization operation.

C. Forgery Localization Module with Multi-level Feature Fusion

Previously, the forgery localization module usually had a single structure or could not adequately integrate the feature

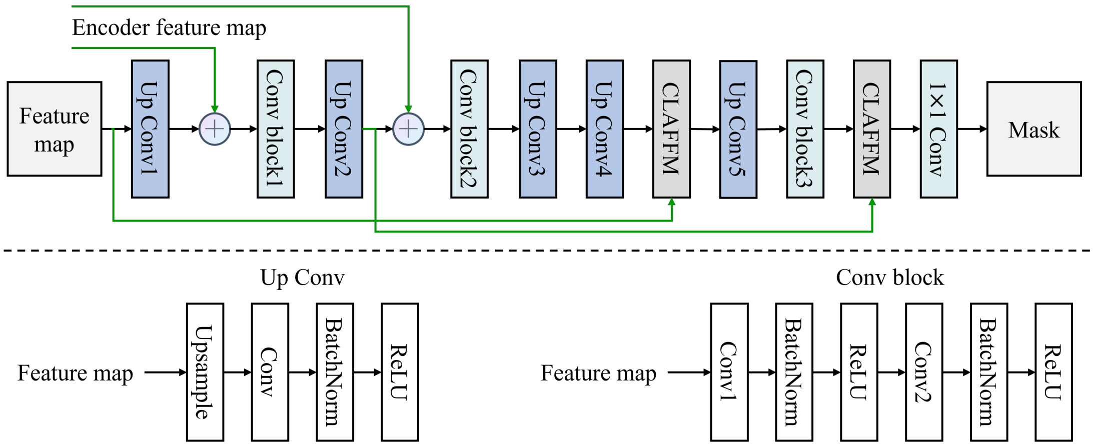


Fig. 4. The detailed architecture of the proposed forgery localization module (FLM).

information of each level, resulting in unsatisfactory outcomes in the forgery region localization task. Therefore, we propose a forgery localization module based on multi-level feature fusion, which made full use of the feature extractor and the feature information of each level of the forgery locator itself, resulting in improved forgery localization effect and detector generality. As shown in Fig. 4, our forgery localization module with multi-level feature fusion receives as input the feature f_i of various block layers of the encoder. Following the up-sampling stage of the forgery locator, the features from different layers of the encoder are coupled with the current features via channels, and more convolution operations are conducted on them to feed into the next up-sampling block. We obtain the forgery localization mask $M^{prd} \in R^{1 \times H \times W}$ for the final output prediction after several layers of upsampling.

TABLE I. Cross-dataset Evaluations on Celeb-DF and DFD.

Methods	FF++(C23)	Celeb-DF	DFD
	AUC	AUC	AUC
Xception [26]	99.09	65.27	87.86
TI2Net [27]	**99.95**	68.22	72.03
FRLM [28]	99.50	70.58	68.17
F3Net [20]	98.10	71.21	86.10
DMGTN [29]	99.80	72.30	—
Face-X-ray [30]	87.40	74.20	85.60
MLDG [31]	98.99	74.56	88.14
GFF [32]	98.36	75.31	85.51
SFDG [33]	99.53	75.83	88.00
SOLA [34]	99.25	76.02	—
MultiAtt [35]	99.27	76.65	87.58
BiG-Arts [36]	99.39	77.04	89.92
LTW [37]	99.17	77.14	88.56
FAAFF [38]	99.27	77.59	—
Local-Relation [17]	99.46	78.26	89.24
DCL [25]	99.30	**82.30**	91.66
Ours	98.54	80.56	**96.01**

Skip connection is a commonly used feature fusion technique in neural networks which can aggregate different levels of semantic information. In addition to combining semantic features at various levels in the feature extractor, we additionally introduce a novel attentional fusion module for combining localization feature information at various levels in the forged localization module. Specifically, our proposed cross-level attentional feature fusion module does not only aggregate features. It is able to obtain a set of adaptive learning weights by adaptively learning the attention of low-level semantic features on the channel. These weights are subsequently utilized to highlight the channel attention of high-level semantic features. This attention technique strengthens the robust local similarity information learned by high-level semantic features while still preserving the local similarity details that low-level semantic features value. Its structure is shown in Fig. 3. For feature A, get previous feature B, first adopt adaptive pooling processing, then perform 4×4 convolution operation and activate it. After that, perform a 1×1 convolution, Sigmoid activate to obtain the attention weight of the same number of channels as the current feature A and output a new feature after multiplying with the feature A. With a very tiny amount of calculation, this module may bring the previous feature's attention information to the present feature and increase the positioning accuracy of the fusion feature based on the channel attention level. The forgery localization loss is defined as follows:

$$L_{LOC} = \frac{1}{N} \sum_{i=1}^{N} BCE \left(M_i^{prd}, M_i \right), \tag{6}$$

where M^{prd} is the prediction mask output by the multi-level forgery localization module, and M is the ground-truth mask.

D. Classifier

In addition to the modules mentioned above, we must additionally include a classifier to receive feature input and verify the image's authenticity. Specifically, the features that the decoder outputs are average pooled and input to a fully connected network classifier for classification. For classifier predictions, we use BCE loss for supervised training:

$$L_{CLS} = \frac{1}{N} \sum_{i=1}^{N} BCE \left(y_i^{'}, y_i \right), \tag{7}$$

where $y^{'} \in [0,1]$ denotes the label of the input image, $y \in \{0,1\}$ denotes the prediction of the classifier. The overall loss

TABLE II. Results of in-dataset Evaluations on FF++ C23 and C40

Methods	**FF++(C23)**		**FF++(C40)**		**Avg**	
	Acc	**AUC**	**Acc**	**AUC**	**Acc**	**AUC**
Face-X-ray [30]	—	87.40	—	61.60	—	74.5
MesoNet [5]	83.10	84.30	70.47	72.62	76.79	78.46
Multi-task[6]	85.65	85.43	81.30	75.59	83.48	80.51
Xception-ELA [39]	93.86	94.80	79.63	82.90	86.75	88.85
SPSL [40]	91.50	95.32	81.57	82.82	86.54	89.07
CFFs [11]	—	97.21	—	86.56	—	91.89
M2TR [41]	91.86	96.75	83.89	87.15	87.88	91.95
Xception [26]	95.73	96.30	86.86	89.30	91.30	92.80
Two-branch [42]	**96.43**	98.70	86.34	86.59	**91.39**	92.65
HFI-Net [43]	91.87	97.07	58.69	88.40	88.78	92.74
RFM [44]	95.69	**98.79**	87.06	89.83	91.38	94.31
FST-Matching [45]	94.05	98.27	**87.38**	90.44	90.72	94.36
Ours	92.84	98.54	81.13	**91.93**	86.99	**95.24**

TABLE III. Results of Cross-dateset Evaluations on FF++ (AUC)

Methods	**Train**	**DF**	**F2F**	**FS**	**NT**	**Avg**
FDFL [21]	**DF**	98.91	58.90	**66.87**	63.61	72.07
MultiAtt [35]		99.92	75.23	40.61	71.08	71.71
GFF [32]		99.87	**76.89**	47.21	72.88	74.21
Ours		**99.99**	73.06	51.86	**76.09**	**75.25**
FDFL [21]	**F2F**	67.55	93.06	55.35	66.66	70.66
MultiAtt [35]		86.15	99.13	60.14	64.59	77.50
GFF [32]		**89.23**	99.10	**61.30**	64.77	**78.60**
Ours		77.70	**99.24**	60.13	**71.52**	77.15
FDFL [21]	**FS**	**75.90**	54.64	98.37	49.72	69.66
MultiAtt [35]		64.13	66.39	99.67	50.10	70.07
GFF [32]		70.21	68.72	**99.85**	49.91	72.17
Ours		70.24	**70.67**	99.65	**54.82**	**73.85**
FDFL [21]	**NT**	79.09	74.21	53.99	88.54	73.96
MultiAtt [35]		87.23	48.22	**75.33**	98.66	77.36
GFF [32]		**88.49**	49.81	74.31	**98.77**	77.85
Ours		83.91	**79.37**	56.64	97.27	**79.30**

function is as follows:

$$L = L_{CLS} + \alpha L_{LS} + \beta L_{LOC}, \tag{8}$$

where α and β are loss weights and range between [0, 1].

IV. Experiments

A. Experimental Settings

1) Datasets: We conduct experiments on several face forery datasets FaceForensics++ (FF++), Celeb-DF-V2, and Deepfake Detection Dataset (DFD) [46]. FF++ is a large-scale public face forgery dataset that contains 1,000 real videos and 4,000 forged videos created using 4 forgery algorithms, including Deepfakes (DF), Face2Face (F2F), FaceSwap (FS), and NeuralTextures (NT). Additionally, FF++ has three compression levels: original version (Raw), high-quality (C23), and low-quality (C40). CelebDF-V2 is a more challenging dataset consists of 569 original videos and 5,639 forged videos. DFD is a large dataset containing 363 real videos and 3,068 forged videos in various scenarios.

2) Implementation details: All faces are detected, cropped, normalized and resized to 224×224 pixels. We use pre-trained Xception as the backbone. All experiments use Adam optimizer with the learning rate set to 1e-4. The batch size is 32, and each epoch has 200 iterations. For the metrics, we utilize binary classification accuracy (Acc.) and area under the curve (AUC.) as the metrics to evaluate model performances.

B. Evaluations

1) In-dataset evaluations: We first evaluate the in-dataset effectiveness of our approach on the FF++, where the network is trained and tested on the same dataset. We only use C23 and C40 versions of FF++ since detecting compressed forged faces is more challenging. We also compare with some state-of-the-art (SOTA) approaches. The results are presented in Table II. We can observe that the proposed approach has promising performances compared to the SOTA approaches when faced with highly compressed forgery faces. The AUC on C23 is close to SOTA [44], while the AUC on C40 exceeds SOTA [45] by 1.49%. This result indicates our method has excellent detection potential. Our proposed method is distinct

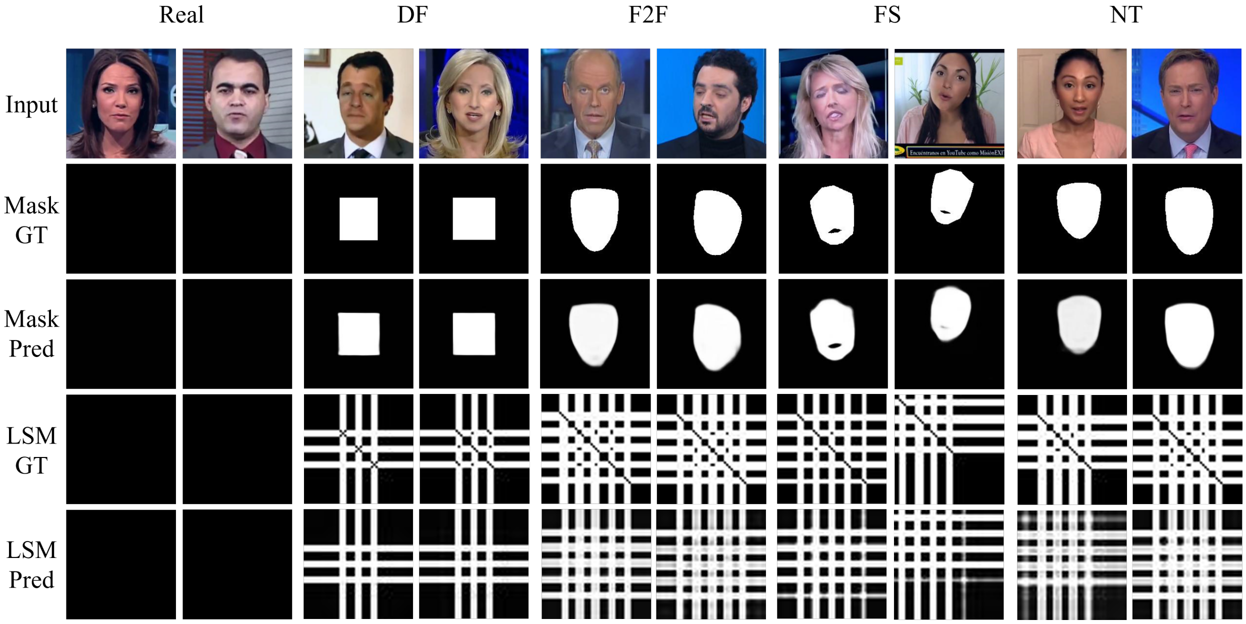


Fig. 5. Visualization of our approach on the FF++. From top to bottom are the original faces, ground-truth of pixel-level mask (Mask GT), predictions of face forgery localization (Mask Pred), ground-truth of local similarity map (LSM GT), and predictions of local similarity map (LSM pred).

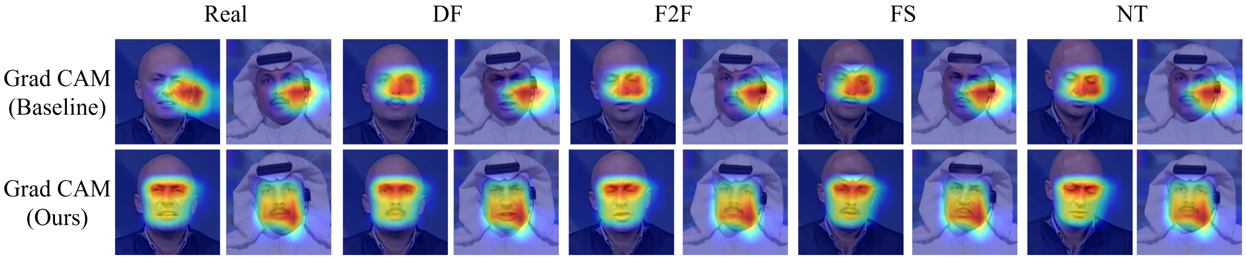


Fig. 6. Visualization of the heatmaps extracted by grad CAM.

from the prior deep learning detection method since it focuses on the fine-grained inconsistency that is shared by various forgery faces rather than just learning the distribution of real and forgery faces, which can effectively improve detection accuracy.

Although we do not achieve the best performance in each setting, we have a clear advantage in high compression scenario (C40) measured by AUC. Moreover, we achieve the best averaged AUC when facing both levels of compression.

2) Cross-dataset evaluations: As we all know, there are countless face manipulation techniques in real scenes, but the face manipulation techniques contained in the samples for training detection models are limited. Therefore, the generalization of detection models based on different manipulation methods is of great practical significance. Cross-dataset evaluations directly reflect the generalization of detectors. Table III presents the cross-dataset evaluation results on FF++(C23). We use only one subset of FF++ for training while the remaining 3 subsets for testing. Our approach achieves the highest average AUC across the three training scenarios. We further evaluate the generalizability of our approach on other datasets. As shown in Table I, we train the model on FF++ (C23) and test it on Celeb-DF and DFD. We can observe that our approach can outperform most recent SOTAs by 2% to 10.00% in terms of AUC while maintaining a promising in-dataset performance, with a gap of less than 1.00% on FF++(C23) compared with SOTAs. On DFD, our approach achieves 96.01% AUC which is 4.35% higher than [25] and nearly 10% better than other approaches on average. The results demonstrate that the proposed learnable local similarity can significantly improve generalization capabilities across different forgeries.

TABLE IV. ABALATION STUDIES ON LLSM AND FLM. THE MODEL IS TRAINED ON FF++(C23) AND EVALUATED ON CELEB-DF (AUC)

Baseline	LLSM	FLM	Celeb-DF
✓			72.15
✓	✓		75.11
✓		✓	76.67
✓	✓	✓	**80.56**

C. Ablation Study

To demonstrate the effectiveness of each module of our approach, we conduct the following ablation studies: 1) Baseline (Xception) without any of the proposed modules; 2) Baseline with the proposed learnable local similarity module; 3) Baseline for the proposed forgery localization module with multi-level feature fusion; 4) The complete network. We present the cross-dataset results trained on FF++ (C23) in Table IV. We can observe that the utilization of LLSM leads to a 2.96% improvement in the performance on Celeb-DF, which confirms the effectiveness of local similarity learning in enhancing generalization. The proposed framework, when incorporated with FLM, demonstrated a 4.52% improvement in performance, affirming its superiority. It is noteworthy that the combined use of FLM and LLSM results in a remarkable performance improvement.

D. Visualization

To provide further evidence of our approach's effectiveness, we evaluate the visualization results using the FF++. As shown in Fig. 5, we compare the forgery localization results and pixel level ground-truth of four different forgery types. The results demonstrate that our approach is capable of performing high-precision forgery localization. Additionally, we present the predictions of LLSM and the corresponding ground-truth similarity map. For real faces, given that the local features share the same source, the similarity maps do not exhibit any abnormal patterns. However, the similarity map of the forged face reveals clear abnormal patterns, which differ depending on the forged region. The LLSM can also achieve accurate predictions on the local similarity patterns. This provides further evidence of the effectiveness of our proposed approach.

We also visualize the heatmaps extracted using Grad CAM [47] to demonstrate the effectiveness of the proposed approach. In these attention maps, the warmer color indicates the areas more significant for predictions or localization. As shown in Fig. 6, we observe that the baseline model is not accurate in identifying the manipulated region, whereas our approach successfully directs the network's attention to the forged facial region.

V. Conclusion

In this paper, we propose a dual-task approach that achieves generalized face forgery detection and accurate forgery localization. Our approach takes advantage of the feature similarity between the internal parts of the forged image. The learnable local similarity module successfully enhances the difference traces between real and forged features and improves the generalization of the model. Furthermore, from a multi-tasking learning view, we present a forgery localization module with cross-level attentional feature fusion strategy, which improves the detection capability even further. We conduct extensive experiments, and the results fully demonstrate the effectiveness of our approach. However, our suggested learnable local similarity module relies on fine-grained local feature calculation, which requires more computational overhead and has feature size restrictions. Moreover, our approach has limited robustness for very low-quality faces. In further studies, we will consider calculating local Inconsistencies for a few local features instead of the entire feature to reduce computational overhead. Additionally, we will design novel image enhancement algorithms to improve the robustness of the detection model.

In the future, exploring the connections between local inconsistencies, identity inconsistencies, and inter-frame inconsistencies may further improve forged face detection performance.

References

[1] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[2] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[3] I. Korshunova, W. Shi, J. Dambre, and L. Theis, "Fast face-swap using convolutional neural networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3677–3685.

[4] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2387–2395.

[5] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in *2018 IEEE international workshop on information forensics and security (WIFS)*. IEEE, 2018, pp. 1–7.

[6] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos," in *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2019, pp. 1–8.

[7] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using capsule networks to detect forged images and videos," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2307–2311.

[8] P. He, H. Li, and H. Wang, "Detection of fake images via the ensemble of deep representations from multi color spaces," in *2019 IEEE international conference on image processing (ICIP)*. IEEE, 2019, pp. 2299–2303.

[9] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*. IEEE, 2019, pp. 83–92.

[10] Z. Liu, X. Qi, and P. H. Torr, "Global texture enhancement for fake face detection in the wild," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8060–8069.

[11] P. Yu, J. Fei, Z. Xia, Z. Zhou, and J. Weng, "Improving generalization by commonality learning in face forgery detection," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 547–558, 2022.

[12] L. Chen, Y. Zhang, Y. Song, L. Liu, and J. Wang, "Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 18 710–18 719.

[13] X. Dong, J. Bao, D. Chen, T. Zhang, W. Zhang, N. Yu, D. Chen, F. Wen, and B. Guo, "Protecting celebrities from deepfake with identity consistency transformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9468–9478.

[14] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, "Protecting world leaders against deep fakes." in *CVPR workshops*, vol. 1, 2019, p. 38.

[15] Z. Shang, H. Xie, Z. Zha, L. Yu, Y. Li, and Y. Zhang, "Prrnet: Pixel-region relation network for face forgery detection," *Pattern Recognition*, vol. 116, p. 107950, 2021.

[16] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning self-consistency for deepfake detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15 023–15 033.

[17] S. Chen, T. Yao, Y. Chen, S. Ding, J. Li, and R. Ji, "Local relation learning for face forgery detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 1081–1088.

[18] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: automatically replacing faces in photographs," in *ACM SIGGRAPH 2008 papers*, 2008, pp. 1–8.

[19] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking," in *2018 IEEE international workshop on information forensics and security (WIFS)*. IEEE, 2018, pp. 1–7.

[20] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII*. Springer, 2020, pp. 86–103.

[21] J. Li, H. Xie, J. Li, Z. Wang, and Y. Zhang, "Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 6458–6467.

[22] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)*. IEEE, 2018, pp. 1–6.

[23] Z. Gu, Y. Chen, T. Yao, S. Ding, J. Li, and L. Ma, "Delving into the local: Dynamic inconsistency learning for deepfake video detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2022, pp. 744–752.

[24] A. Haliassos, K. Vougioukas, S. Petridis, and M. Pantic, "Lips don't lie: A generalisable and robust approach to face forgery detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 5039–5049.

[25] K. Sun, T. Yao, S. Chen, S. Ding, J. Li, and R. Ji, "Dual contrastive learning for general face forgery detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2022, pp. 2316–2324.

[26] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1–11.

[27] B. Liu, B. Liu, M. Ding, T. Zhu, and X. Yu, "Ti2net: Temporal identity inconsistency network for deepfake detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4691–4700.

[28] C. Miao, Q. Chu, W. Li, S. Li, Z. Tan, W. Zhuang, and N. Yu, "Learning forgery region-aware and id-independent features for face manipulation detection," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 1, pp. 71–84, 2021.

[29] B. Liang, Z. Wang, B. Huang, Q. Zou, Q. Wang, and J. Liang, "Depth map guided triplet network for deepfake face detection," *Neural Networks*, vol. 159, pp. 34–42, 2023.

[30] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo, "Face x-ray for more general face forgery detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5001–5010.

[31] D. Li, Y. Yang, Y.-Z. Song, and T. Hospedales, "Learning to generalize: Meta-learning for domain generalization," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.

[32] Y. Luo, Y. Zhang, J. Yan, and W. Liu, "Generalizing face forgery detection with high-frequency features," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 16 317–16 326.

[33] Y. Wang, K. Yu, C. Chen, X. Hu, and S. Peng, "Dynamic graph learning with content-guided spatial-frequency relation reasoning for deepfake detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7278–7287.

[34] J. Fei, Y. Dai, P. Yu, T. Shen, Z. Xia, and J. Weng, "Learning second order local anomaly for general face forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 270–20 280.

[35] H. Zhao, W. Zhou, D. Chen, T. Wei, W. Zhang, and N. Yu, "Multi-attentional deepfake detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2185–2194.

[36] H. Chen, Y. Li, D. Lin, B. Li, and J. Wu, "Watching the big artifacts: Exposing deepfake videos via bi-granularity artifacts," *Pattern Recognition*, vol. 135, p. 109179, 2023.

[37] K. Sun, H. Liu, Q. Ye, Y. Gao, J. Liu, L. Shao, and R. Ji, "Domain general face forgery detection by learning to weight," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, 2021, pp. 2638–2646.

[38] C. Tian, Z. Luo, G. Shi, and S. Li, "Frequency-aware attentional feature fusion for deepfake detection," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[39] T. S. Gunawan, S. A. M. Hanafiah, M. Kartiwi, N. Ismail, N. F. Za'bah, and A. N. Nordin, "Development of photo forensics algorithm by detecting photoshop manipulation using error level analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 7, no. 1, pp. 131–137, 2017.

[40] H. Liu, X. Li, W. Zhou, Y. Chen, Y. He, H. Xue, W. Zhang, and N. Yu, "Spatial-phase shallow learning: rethinking face forgery detection in frequency domain," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 772–781.

[41] J. Wang, Z. Wu, W. Ouyang, X. Han, J. Chen, Y.-G. Jiang, and S.-N. Li, "M2tr: Multi-modal multi-scale transformers for deepfake detection," in *Proceedings of the 2022 International Conference on Multimedia Retrieval*, 2022, pp. 615–623.

[42] I. Masi, A. Killekar, R. M. Mascarenhas, S. P. Gurudatt, and W. AbdAlmageed, "Two-branch recurrent network for isolating deepfakes in videos," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*. Springer, 2020, pp. 667–684.

[43] C. Miao, Z. Tan, Q. Chu, N. Yu, and G. Guo, "Hierarchical frequency-assisted interactive networks for face manipulation detection," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3008–3021, 2022.

[44] C. Wang and W. Deng, "Representative forgery mining for fake face detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14 923–14 932.

[45] S. Dong, J. Wang, J. Liang, H. Fan, and R. Ji, "Explaining deepfake detection by analysing image matching," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIV*. Springer, 2022, pp. 18–35.

[46] N. Dufour and A. Gully, "Contributing data to deepfake detection research," *Google AI Blog*, vol. 1, no. 3, 2019.

[47] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

CLFM: Contrastive Learning and Filter-attention Mechanism for Joint Relation Extraction

Zhiyuan Wang, Chuyuan Wei*, Jinzhe Li, Lei Zhang, Cheng Lv
School of Electrical and Information Engineering,
Beijing University of Civil Engineering and Architecture
Beijing 100044 China

Abstract—**Relation extraction is a fundamental task in natural language processing, which involves extracting structured information from textual data. Despite the success of joint methods in recent years, most of them still have the propagation of cascade errors. Specifically, the error in former step will be accumulated into the final combined triples. Meanwhile, these methods also encounter another challenges related to insufficient interaction between subtasks. To alleviate these issues, this paper proposes a novel joint relation extraction model that integrates a contrastive learning approach and a filter-attention mechanism. The proposed model incorporates a potential relation decoder that utilizes contrastive learning to reduce error propagation and enhance the accuracy of relation classification, particularly in scenarios involving multiple relationships. It also includes a relation-specific sequence tagging decoder that employs a filter-attention mechanism to highlight more informative features, alongside an auxiliary matrix that amalgamates information related to entity pairs. Extensive experiments are conducted on two public datasets and the results demonstrate that this approach outperforms other models with the same structure in recall and F1. Moreover, experiments show that both the contrastive learning strategy and the proposed filter-attention mechanism work well.**

Keywords—Natural language processing; relation extraction; attention mechanism; contrastive learning; multi-task learning

I. Introduction

Relation extraction intend to extract pairs of entity and correlative relations in the form of $< subject, relation, object >$ from the given unstructured texts. The extracted information provides a supplement to many natural language processing(NLP) tasks, such as text summarization [1] knowledge graph construction [2] and question answering [3].

Conventionally, existing methods mainly include pipeline methods and joint methods. Pipeline works [4], [5], [6], [7] traditionally treat the task as two independent subtasks: named entity recognition (NER) and relation extraction (RE). While these approaches are straightforward and adaptable, they overlook the inherent connection between NER and RE, making them prone to error propagation due to the conventional order of subtasks. For this reason, most recent studies focus on joint methods [8], [9], [10]. Current joint models, as evidenced by the work of Cabot et al. in REBEL [11] and Zheng et al. in PRGC [12], have demonstrated remarkable efficiency while achieving outstanding performance. However, most of them first identify entities then find corresponding relationships from all predefined relationships, which faces the problem of relational redundancy and cause unnecessary calculations. Fig. 1 shows the difference between entity first and relation first methods. The entity first method always recognizes possible entity pairs and then matches them against all predefined relationships, which causes redundancy in relationships and introduces unnecessary computation into the model. In contrast, the relation first approach avoids this problem well by first recognizing the relations present in the sentence. Another problem is that some methods only perform simple interactive behaviors between sequence and potential relation representations like concatenating. The operation could carry information unrelated to the task and cannot fully utilize useful mutual information. In addition to this, most of the models suffer from General issue of error propagation. The error in former step will be accumulated in the final triplets.

To alleviate error propagation issue and enhance the interaction between two subtasks, this work makes use of a contrastive learning strategy and proposes a filter-attention mechanism for RE and relation-specific NER(CLFM), respectively. Specifically, the contrastive learning employ the R-drop [13] idea, which has been used in supervised image classification for computer vision tasks. The contrastive learning strategy brings a new constraint to the part of potential relation classification rather than just relying on cross-entropy loss. This can lead to more accurate classification results, especially in scenarios with multiple relationships. The potential relation classification module employs the R-drop idea, which has been used in supervised image classification for computer vision tasks. This strategy brings a new constraint to the part of potential relation classification rather than just relying on cross-entropy loss. This can lead to more accurate classification results, especially in scenarios with multiple relationships, thus the problem of error propagation can be mitigated. For the relation-specific NER task, a novel filter-attention mechanism based on the attention mechanism [14] is proposed. Unlike prior works that simply concatenate or add sentence and relation representations, task interaction in this method is achieved in two ways: Initially, attention scores are computed for sentence and relation representations to signify the token relevance concerning the current relation. Next, low scores indicating weak correlations within the score matrix are removed, followed by the concatenation of particular representations to form the input for the relation-specific Named Entity Recognition (NER) task. Regarding NER, it is perceived as a conventional sequence tagging task for acquiring potential entity pairs. This process strengthens the interaction between

*Corresponding authors

the two subtasks and eliminates inconsequential information across distinct relationships. While executing the relation classification task, simultaneous computation of the auxiliary matrix occurs, which is subsequently utilized to determine the final triplets.

CLFM is mainly composed of three parts: All possible relations in the input sentence are identified in the first part; The second part is the identification of all possible head entities and tail entities under the specific relationships that have been extracted earlier; Finally, the model uses an auxiliary matrix called subject-object alignment step to help select the final triplets from extracted entity pairs and relationships. Since the order of the subtasks, the proposed end-to-end method can also solve the problem of redundant relationships.

The approach has been tested on two widely-used public datasets, namely, NYT [15] and WebNLG [16]. Experimental results indicate the model's performance is on par with state-of-the-art methods on these benchmark datasets. In summary, the paper's key contributions are as follows:

1. A relation first end-to-end framework is introduced by this work, along with the design of three components pertaining to the subtasks. These components effectively tackle issues related to redundant relations and enhance the interaction between subtasks.

2. This work incorporates a contrastive learning strategy into relation classification to introduce a new constraint that enhances classification accuracy and mitigates error propagation.

3. An innovative filter-attention mechanism is presented in this work, with the objective of fostering deeper interaction between the two subtasks by proficiently filtering out irrelevant task-specific information. Extensive experimentation on various public benchmarks demonstrates results that exhibit performance comparable to the baseline, especially in scenarios involving multiple relationships.

This paper is structured as follows: Section II is related work of recent years in relation extraction; Section III are detailed description of the each model part; Section IV are experimental results compared to other baseline methods and ablation experiments; Section V are conclusion and future work.

II. Related Work

The traditional incipient relational triplet extraction methods, such as those proposed by Zelenko et al. [4] and Chan et al. [5], adopted a pipeline framework that divided the entire task into two separate subtasks: entity identification and relation classification. However, these approaches were prone to error propagation problems and ignored the fact that these subtasks are interactive. Therefore, later works began to extract entities and relations jointly using a single model, such as feature-based models [17], [18], [19]. These models rely on various external NLP tools and complicated manual operations, making them heavily dependent on the accuracy of these external tools. Although these models are very representative, they have limitations in terms of scalability and efficiency. In the past few years, joint methods based on neural networks have become a major research focus. This section is presented in three subsections: entity first methods, relation first methods and other methods.

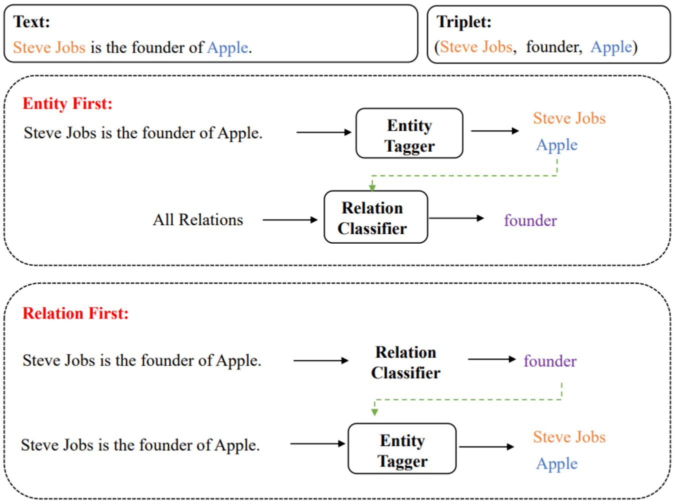

Text:
Steve Jobs is the founder of Apple.
Triplet:
(Steve Jobs, founder, Apple)
Entity First:
Steve Jobs is the founder of Apple.
Entity Tagger
Steve Jobs
Apple
All Relations
Relation Classifier
founder
Relation First:
Steve Jobs is the founder of Apple.
Relation Classifier
founder
Steve Jobs is the founder of Apple.
Entity Tagger
Steve Jobs
Apple

Fig. 1. The difference between entity first and relation first approaches.

A. Entity First

Zheng et al. [20] first proposed a novel joint model based on a tagging scheme, which transformed the relation extraction task into a sequence labeling task and applied Long Short-Term Memory network to learn long-term dependencies. However, the model has no ability to extract overlapping triplets. To address the overlapping issue, Yu et al. [21] proposed a method that extracts head entities in the first step, followed by all correlative tail entities and relations using decoding strategies. In addition, A unified joint extraction annotation framework was designed by Wei et al. [22], capable of achieving single-stage joint extraction while addressing exposure bias and the intricate issue of overlapping. In the work by Wang et al. [23], a one-stage approach was presented to simultaneously extract entities and overlapping relations. This approach effectively narrows the discrepancy between training and inference stages. Specifically, they formulated joint extraction as a token pair linking problem and introduced a novel handshaking tagging scheme that aligns the boundary tokens of entity pairs under each relation type. Shang et al. [24] first generated candidate entities by enumerating sequences of tokens in a sentence, and they converted the extraction task into a linking problem on a head-to-tail bipartite graph that could directly extract all triplets in a single step. Nayak et al. [25] proposed a pointer network-based decoding approach where an entire tuple is generated at every time step and achieved significantly higher F1 scores

B. Relation First

All of the approaches mentioned, regardless of being single-stage or not, suffer from relational redundancy issue [20], [22], [23]. As a result, a new model structure for extracting sequences has emerged. These methods typically involve performing relation classification first, which only preserves related relations and not all redundant relations in the input sentence. However, Yuan et al. [8] proposed a gating mechanism to obtain relation-specific sentence representation, which can be used for sequence tagging tasks and can provide a fine-grained representation. Nevertheless, this mechanism is unable to address the problem of subject-object overlapping. To address this issue, Ma et al. [26] proposed a cascade dual-decoder approach to extract overlapping relational triples. This approach utilizes relevant information of relations and subjects as auxiliary information for subjects and objects recognition, respectively. However, the approach still has poor generalization due to insufficient interaction and a span-based extraction strategy. Zheng et al. [12] decomposed the entire task into three subtasks: relation judgement, entity extraction, and subject-object alignment. They designed a low complexity global correspondence matrix to align the subject and object. Despite achieving success, this approach lacks deep interaction between relation classification and entity recognition.

C. Other Method

Shang et al. [27] proposed a one-step and one-module model that consists of a scoring-based classifier and a relation-specific horns tagging strategy. Zhao et al. [28] tackled the task of relation extraction using heterogeneous graph neural networks. Their approach involves modeling relations and words as nodes on a graph and iteratively fusing the two types of semantic nodes using a message passing mechanism to obtain node representation. This approach leverages the graph structure and takes into account the contextual information of both relations and words. Ning et al. [29] considered the extraction task based on the table-filling method as a target detection task and proposed a single-stage target detection framework, which combined with the auxiliary global relational triplets region detection to ensure the region information can be fully utilized. Ye et al. [30] employed the different strategies for NER and RE by using solid and levitated markers of neighboring spans inside the same sample.

III. Proposed Method

A. Problem Formulation

Given the input sentence $S = \{x_1, x_2, \ldots, x_n\}$ with n tokens, the task goal is to extract all possible relational triplets such as $\{T = (s, r, o) \mid s, o \in E, r \in R\}$, where E and R are entity and relation sets respectively, a triplet T represents a pair of entity and a relation between them contained in sentence S.

As shown in Fig. 2, given an input sentence S, the encoder starts modeling its text semantics. The potential relation decoder, with a contrastive learning strategy, detects all possible relations $r \in R$ based on the text semantics. For each detected relation r, the filter-attention mechanism computes and filters the low weights of each input token to get relation-specific sentence representation as the input of NER. The relation-specific entity decoder extracts the corresponding head and tail entities by using a sequence tagging scheme. Finally, the model obtains the final triplets T with the aid of an auxiliary matrix M which were generated at the same stage as the potential relation prediction.

B. Model Encoder

This approach employs a pre-trained language model called BERT [31] for a fair comparison, which is widely used to encode sentences and capture the semantics of text. The output of the model encoder is $H_{enc} = \{h_1, h_2, \ldots, h_n \mid h_i \in \mathbb{R}^{d\times 1}\}$, where n is the number of tokens, and d is the dimension of the embedding. It can also use other pre-trained language models, such as RoBERTa [32] and so on.

C. Relation Classification

The relation classification component is illustrated in Fig. 2. It is important to note that not all sentences contain all predefined relations. Therefore, CLFM starts by identifying potential relations, which helps to reduce redundant relationships in the current text. To complete the operation, average pooling and a fully connected layer are employed. Given the embedding $\mathbf{h} \in \mathbb{R}^{n\times d}$ of the input sentence, where n is the number of tokens, each element of the relation classification component is obtained as follows:

$$\begin{aligned} h^{pool} &= \varphi(\mathbf{h}) \in \mathbb{R}^{d\times 1} \\ P^{pot} &= \sigma(W_r h^{pool} + b_r) \end{aligned} \tag{1}$$

Where φ denotes the average pooling operation [33] and σ is the sigmoid activation function, $W_r \in \mathbb{R}^{d\times 1}$ is a trainable weight and b_r is a parameter.

Unlike previous works [21], [26], [12], which treat relation classification as a simple binary classification task, this approach incorporates a contrastive learning strategy to add a new constraint to the result of the relation classification task. Inspired by [13], CLFM employs the idea of R-drop to impose a new constraint on the result of the potential relation classification task. Specifically, it computes the Kullback-Leibler divergence of the classification results as a part of the relational loss by running the sentence embedding $\mathbf{h}$ through the sentence classifier twice. Since the classifier contains a dropout operation, two results for the same input may be different, which can increase the robustness of the classifier by continuous training. For classification results, if the probability exceeds a threshold λ_1, model allocates the corresponding relation a tag of 1; otherwise, model assigns a tag of 0. The detailed contrastive learning component and potential relation loss are as follows:

$$L_{cl} = \frac{1}{2}(D_{KL}(P_1^{pot} \parallel P_2^{pot}) + D_{KL}(P_2^{pot} \parallel P_1^{pot})) \tag{2}$$

$$\mathcal{L}_{rc} = -\frac{1}{n_r}\sum_{i=1}^{n_r}(y_i \log P^{pot} + (1-y_i)\log\left(1-P^{pot}\right)) \tag{3}$$

$$L_{rel} = \alpha L_{rc} + \beta L_{cl} \tag{4}$$

Where P_1^{pot} and P_2^{pot} are transformed into a predefined relational representation by P^{pot}, $D_{KL}(\|)$ denotes Kullback-Leibler divergence and n_r is the size of predefined relation set. α and β are weights of each sub-loss. Performance might be better by carefully tuning the weight of each sub-loss. The reason why this work takes the average of the two calculations as the result is that Kullback-Leibler divergence is asymmetric. After adding the new constraint, CLFM can better handle of datasets with more relationships.

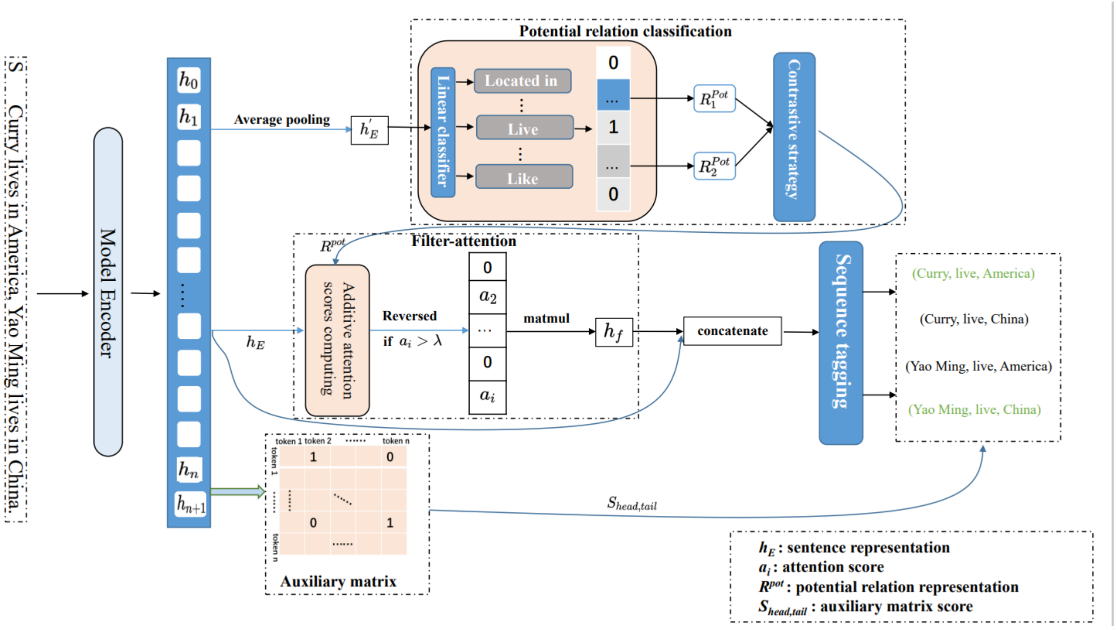


Fig. 2. The overall structure of CLFM. It combines three parts: potential relation classification, sequence labeling and auxiliary matrix.

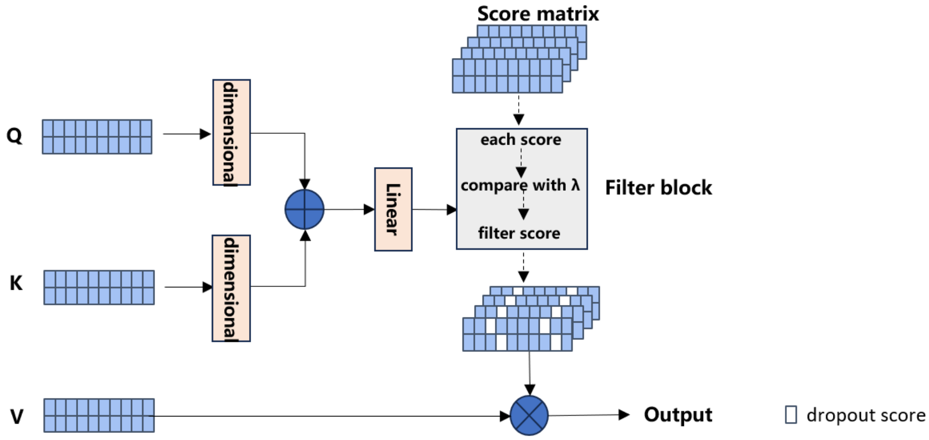


Fig. 3. The process of filter-attention.

D. Filter-attention Mechanism

After obtaining all potential relationships in the current sentence, previous works typically concatenate sentence and relation embeddings or design complex gating mechanism. However, concatenation can lead to the introduction of useless information. Compared to than gating mechanism, attention mechanism is more intuitive and easier to understand and the elements are more closely related. Therefore, this paper designs an attention mechanism with a filtering function based on attention mechanism [14] to retain useful mutual information. Fig. 3 shows the details of filter-attention mechanism. Before starting relation-specific NER, sentence embedding and relation embedding are fed into filter-attention. It utilizes additive attention to capture diverse semantic information across various relationships, as it aligns better with the encoder-decoder structure. After computing attention scores, filter module selects higher scores for retention to obtain a more accurate representation of the input containing information about the current relationship. The filter-attention mechanism component is as follows:

$$\begin{aligned} &S(\mathbf{h}, h_r) = W_v\theta(W_q\mathbf{h} + W_k h_r) \\ &S_{filter}(\mathbf{h}, h_r) = F(Softmax(S(\mathbf{h}, h_r))) \\ &\mu = S_{filter}(\mathbf{h}, h_r) \cdot h_r \end{aligned} \tag{5}$$

Where W_q, $W_k \in \mathbb{R}^{d\times d}$ and $W_v \in \mathbb{R}^{d\times 1}$ are trainable weights, h_r is transformed into a predefined relation representation by P^{pot}. $\theta(\cdot)$ is tanh activation function and $F(\cdot)$ is the filter operation. A threshold λ_{att} is established for the filter mechanism. When scores exceed λ_{att}, the scores in the original matrix are retained; otherwise, they are discarded. After conducting thorough experiments, λ_{att} is assigned a value of 5e-3. The ultimate representation that undergoes the filter-attention mechanism is denoted as $u \in \mathbb{R}^{d\times 1}$.

E. Relation-specific NER

As shown in Fig.2, the model starts the relation-specific entity recognition task after completing the filter-attention operation. CLFM model it as a sequence tagging task because the generalization of span-based extraction methods is poor. To solve common overlapping triplet issues, it perform separate sequence tagging for head entity and tail entity. This strategy can handle issues including EntityPairOverlap (EPO) and SingleEntityOverlap (SEO). The sequence tag set for the head entity is $\{B\text{-}H, I\text{-}H, O\}$, and the sequence tag set for the tail entity is $\{B\text{-}T, I\text{-}T, O\}$. CLFM employ the traditional LSTM-CRF[10] network for sequence tagging, and the detailed formula descriptions are as follows:

$$\begin{aligned} o_{i,j} &= \left[\overrightarrow{\mathrm{LSTM}}\left(h_i;\mu_j\right); \overleftarrow{\mathrm{LSTM}}\left(h_i;\mu_j\right)\right] \\ P_{i,j}^{head} &= \mathrm{CRF}\left(o_{i,j}\right) \end{aligned} \tag{6}$$

Where $(;)$ denotes concatenating operation, h_i is $i\text{-}th$ token representation of sentence S and μ_j is $j\text{-}th$ relation representation after going through filter-attention component. $\mathrm{CRF}(\cdot)$ is Conditional Random Field approach. The head entity formula is only given, tail entity formula is the same as Eq. (6). The loss of whole entity recognition is Eq. (7):

$$\mathcal{L}_{seq} = -\frac{1}{2 \times n \times n_r^{pot}} \sum_{t \in \{head,tail\}} \sum_{j=1}^{n_r^{pot}} \sum_{i=1}^{n} y_{i,j}^{t} \log P_{i,j}^{t} \tag{7}$$

Where n_r^{pot} is the size of potential relation set of sentence S.

F. Auxiliary Matrix and Training Strategy

After all the above operations, CLFM gets all possible head and tail entities that correspond to specific relations. However, if the model directly outputs the outcome, there will be a high probability of obtaining inaccurate triplets. To avoid this situation, following [12], the model computes an auxiliary matrix $M \in \mathbb{R}^{n \times n}$ for the given sentence S with n tokens to denote whether a relationship exists between tokens. This is similar to a pruning operation that increases the limits and make the final triplets more accurate. The value of each element in the matrix is computed as follows:

$$P_{i_{head},j_{tail}} = \sigma\left(W_g\left[h_i^{\text{head}}; h_j^{\text{tail}}\right] + b_g\right) \tag{8}$$

Where h_i^{head} , $h_j^{\text{tail}} \in \mathbb{R}^{d \times 1}$ are the encoded representation of the $i\text{-}th$ token and $j\text{-}th$ token in the input sentence forming a potential pair of head and tail entities. W_g is a trainable weight.

Matrix loss is as follows:

$$\begin{aligned} \mathcal{L}_{\text{matrix}} = -\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (y_{i,j} \log P_{i_{\text{head}},j_{\text{tail}}} \\ + \left(1 - y_{i,j}\right) \log\left(1 - P_{i_{\text{head}},j_{\text{tail}}}\right)) \end{aligned} \tag{9}$$

The total loss is the sum of these three parts:

$$\mathcal{L}_{totalloss} = \gamma_1 L_{rel} + \gamma_2 L_{seq} + \gamma_3 L_{matrix} \tag{10}$$

where $\gamma_1, \gamma_2, \gamma_3$ are adjustable loss weights.

IV. Experiment

A. Datasets and Evaluation Metric

For fair and comprehensive comparison, this work follow [22], [9], and [21] to evaluate CLFM on two widely used public datasets: NYT and WebNLG. The NYT dataset is generated by aligning the relations in Freebase with the New York Times (NYT) corpus and is widely used for remotely supervised relational extraction tasks. It contains 24 relation types. The WebNLG dataset, originally employed for natural language generation tasks, includes 246 relation types. It is worth noting that both datasets have another version: NYT* and WebNLG*, which annotate the last word of entities. Table I shows the statistics for the above datasets.

Two evaluation metrics are employed for expensive experimental studies: Partial Match for NYT* and WebNLG*, where an extracted triple (s, r, o) is regarded as correct only if its relation and the last word of the head entity name and the tail entity name are correct; Exact Match for NYT and WebNLG, where a predicted triple (s, r, o) is regarded as correct only if its relation and the full names of its head and tail entities are all correct.

TABLE I. The Statistics of Datasets

Datasets	Train	Valid	Test	Relations
NYT	56196	5000	5000	24
WebNLG	5019	500	703	171
NYT*	56195	4999	5000	24
WebNLG*	5019	500	703	216

B. Implementation Details

As presented in Fig. 2, CLFM encoder employs the PyToREh version of $BERT_{base}\,(cased)$ English. To ensure equitable comparison, the input sentence length is standardized to a fixed size of 100, and the Adam optimizer [34] is employed with a batch size of 32/6 for the NYT/WebNLG datasets. The learning rate for the BERT encoder is set to 5e-5, while the decoder learning rate is set to 1e-3 to achieve fast convergence. Moreover, the utilization of weight decay [35] at a rate of 0.01 is incorporated. The potential relation decoder threshold and filter-attention threshold are set as 0.5 and 5e-3, respectively.

The experiments are conducted on a server equipped with Intel(R) Xeon(R) Silver 4215 CPU @ 2.50GHz and an NVIDIA Tesla V100 GPU.

C. Baseline Methods

A selection of nine baseline methods has been made for the purpose of comparison. This assortment comprises representative models as well as models featuring analogous structures. CLFM is compared with the following strong baseline models on the NYT and WebNLG datasets. The top six models are representative methods, while the last three have similar structures to CLFM: (1) CasRel[22] (2) TPLinker[23] (3)WDec[25](4) CGT[36] (5) StereoRel[37] (6) RIFRE[28] (7)PRGC[12] (8) RSAN[8] (9) Cascade dual-decoder[26].

TABLE II. PERFORMANCE OF CLFM AND EIGHT COMPARED BASELINES ON NYT AND WEBNLG

Model	NYT			WebNLG			NYT*			WebNLG*		
	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1
CasRel	-	-	-	-	-	-	84.2	83.0	83.6	86.9	80.6	83.7
TPLinker	91.4	92.6	92.0	88.9	84.5	86.7	91.3	92.5	91.9	91.8	92.0	91.9
WDec	88.1	76.1	81.7	88.6	51.3	65.0	94.5	76.2	84.4	-	-	-
CGT	-	-	-	-	-	-	94.7	84.2	89.1	92.9	75.6	83.4
StereoRel	92.0	92.3	92.2	-	-	-	92.0	92.3	92.2	91.6	92.6	92.1
RIFRE	-	-	-	-	-	-	93.6	90.5	92.0	93.3	92.0	92.6
PRGC	93.5	91.9	92.7	89.9	87.2	88.5	93.3	91.9	92.6	94.0	92.1	93.0
RSAN	85.7	83.6	84.6	80.5	83.8	82.1	-	-	-	-	-	-
Cascade dual-decoder	89.9	91.4	90.6	88.0	88.9	88.4	90.2	90.9	90.5	90.3	91.5	90.9
CLFM	93.3	**92.4**	**92.8**	**90.3**	87.9	**89.1**	93.0	92.3	**92.7**	93.9	**92.6**	**93.3**

D. Experimental Results

This section presents experimental results and compares them with other baseline models. It also conduct an in-depth analysis of the results to gain a better understanding of the performance of CLFM in relation to the other methods. Through a comprehensive analysis of the outcomes, valuable insights into the workings of CLFM can be acquired.

1) Overall Results: Table II presents an overall comparison of CLFM with other baselines. CLFM outperforms all of the baselines in terms of F1 scores, and for most cases, precision and recall are also superior. Notably, CLFM exhibits better robustness and generalization on the WebNlG dataset, where a wide range of relations are involved. This success can be attributed to the incorporation of the introduced contrastive learning strategy, which significantly enhances the accuracy of the relation classification decoder. Although RSAN[8] also employs attention calculations through a gate mechanism, CLFM achieves superior performance (with at least an 8% improvement over RSAN) due to its simplicity and commonality. Additionally, the model achieve a 0.6% improvement on the WebNLG dataset compared to PRGC. Regarding the NYT dataset, although CLFM achieves similar performance as PRGC, it is believe that the limited number of relations in the dataset and the already high-quality results may have contributed to the lack of significant improvement.

2) Result Analysis on Different Sentence Types : Following previous works [26], we conduct extensive experiments to verify that our method makes effective in the scenario of overlapping triples on NYT and WebNLG datasets. Table III shows the detailed results on the three overlapping patterns, where Normal is the easiest pattern while EPO and SEO are more difficult to be handled. The experimental findings reveal a consistent and superior performance exhibited by our proposed model across all three overlapping patterns. Noteworthy is the model's exceptional efficacy in handling intricate patterns such as EPO and SEO, where it consistently outperforms the established baseline, PRGC. This substantiates the robust capabilities of our model in effectively addressing the intricate challenges posed by overlapping triplets.

Furthermore, an extensive examination was conducted to extract triples from sentences featuring varying numbers of triplets. The sentences were categorized into five subclasses, each encompassing texts with 1, 2, 3, 4, or $\geq$ 5 triples. The comparative results of the five methods across the different triple categories are depicted in Fig. 4. The figure illustrates that our model consistently achieves the highest F1 scores across most cases in the two datasets, exhibiting remarkable stability as the number of triples increases. Particularly noteworthy is the superior performance of our model in comparison to the leading baseline, PRGC, in the most challenging class ($\geq$ 5) on NYT and WebNLG. This suggests that our model demonstrates enhanced resilience in handling intricate scenarios involving a substantial number of triples.Moreover, our model outperforms others in nearly every subset, irrespective of the number of triples. In summary, these additional experiments substantiate the advantageous features of our model, particularly in complex scenarios, highlighting its robustness and superior performance over existing methods.

TABLE III. F1-SCORE OF SENTENCES WITH DIFFERENT OVERLAPPING TRIPLETS ON NYT AND WEBNLG

Model	NYT			WebNLG		
	Normal	SEO	EPO	Normal	SEO	EPO
WDec	80.3	81.4	86.7	75.5	63.3	67.0
CasRel	84.2	83.0	83.6	86.9	80.6	83.7
PRGC	88.6	93.6	94.1	86.8	89.0	89.8
Cascade dual-decoder	88.2	92.8	92.9	86.2	88.9	88.5
CLFM	**90.9**	**94.4**	**94.7**	**87.6**	**89.6**	**94.1**

3) Explore for Filter Threads: The performance of the filter-attention mechanism is highly influenced by the threshold λ_{att}. To study the impact of threshold changes and find the appropriate value, A series of experiments are carried out. Table IV shows the results of these experiments. Variations were introduced to the threshold within a predefined range, revealing that further enhancements in the experimental outcomes could be achieved by implementing additional threshold adjustments. Based on the current results, A threshold value of 5e-3 was chosen for the filter-attention mechanism. Nevertheless, it is important to highlight that the optimal threshold might exhibit variability contingent on the specific dataset or task. As a consequence, undertaking additional experiments and meticulous refinement of the threshold could be imperative

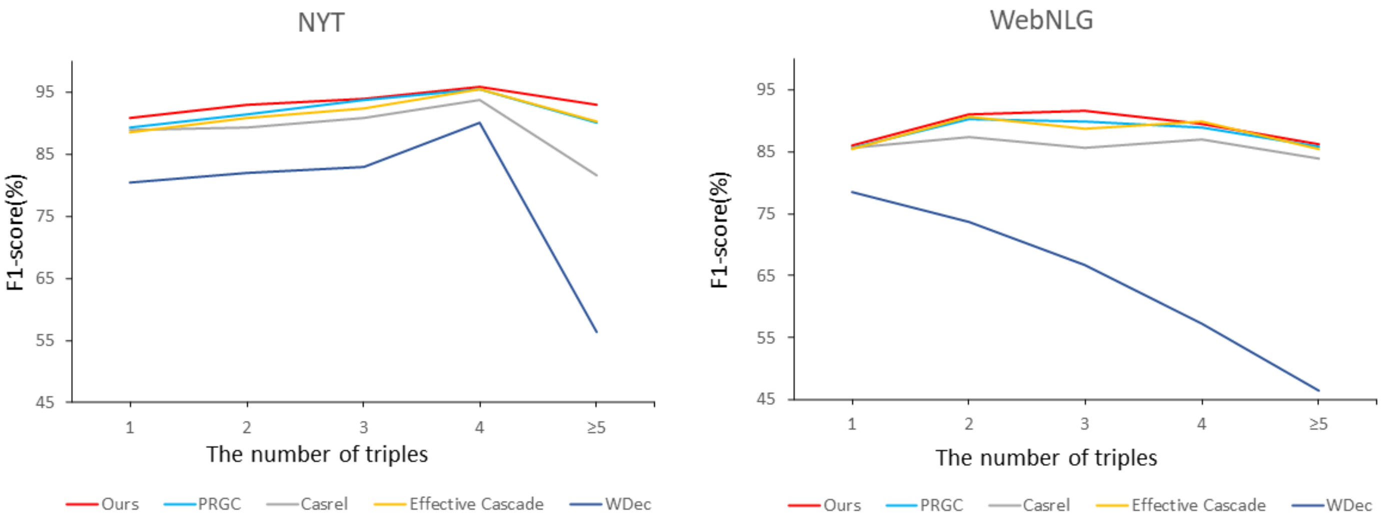


Fig. 4. Results of different sentence types on NYT and WebNLG.

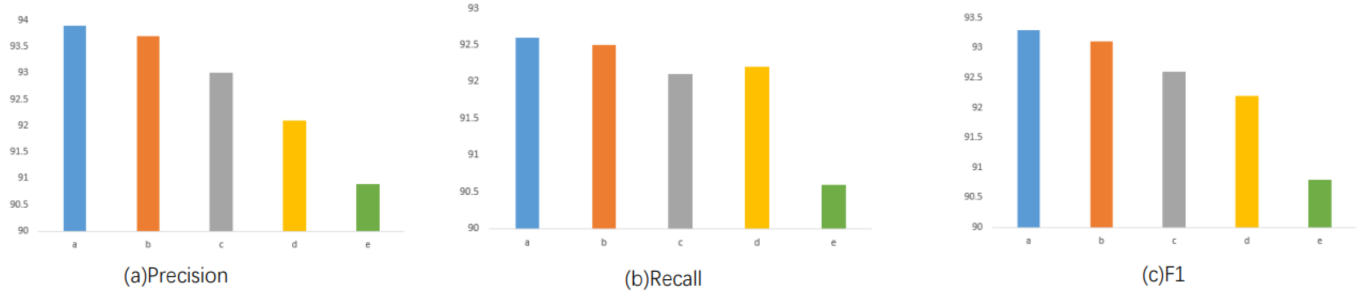


Fig. 5. The results of different combination approaches.

for attaining superior results.

TABLE IV. The Experimental Results of Different Threads on WebNLG*

Thread	F1(WebNLG*)
1e-3	92.8
3e-3	92.4
4e-3	92.0
5e-3	**93.3**
7e-3	92.0

4) Ways of Filter-attention: Furthermore, diverse methodologies for amalgamating the representations acquired via the filter-attention mechanism with the sentence embeddings are explored. To facilitate discourse, the nomenclature adopted encompasses the relation embeddings procured via the filter-attention mechanism, denoted as filter-relation, the sentence embeddings acquired through the filter-attention mechanism, termed as filter-sequence, and the sentence embeddings generated by the BERT encoder, identified as sentence-output. Concatenation function is represented by the $(\cdot, \cdot)$ notation. As illustrated in Fig. 5, five distinct strategies for amalgamating the embeddings are subjected to experimentation on the WebNLG* dataset, namely, (a) sequence-output, filter-relation, (b) filter-sequence, (c) sequence-output, filter-relation + filter-sequence, (d) sequence-output, filter-sequence, (e) filter-sequence, filter-relation. Precision, recall, and F1 score are used to evaluate the performance of each approach. The experiment results demonstrate that the first Combination method, (sequence-output, filter-relation), outperforms the other four combinations and is currently the best approach. Additionally, this validation substantiates the rationale behind the inception of the filter-attention mechanism.

5) Ablation Study: This section examines the contributions of different modules components in CLFM, using the best performing model on the WebNLG* dataset. Initially, the contrastive learning component is removed. Subsequently, an exploration is conducted into the influence of the filter-attention mechanism. This entails retaining the attention mechanism while discarding the filter function, as well as removing both components. Table V shows the results. It can observe that contrastive strategy and filter-attention mechanism can improve the performance of the model. The results demonstrate that contrastive learning can enhance the accuracy of the relation decoder when dealing with multiple relations, while filter-attention can improve the quality of the interactive representations between two subtasks.

TABLE V. ABLATION STUDY RESULTS ON WEBNLG* TEST SET

Method	Pre.	Rec.	F1
CLFM	**93.9**	**92.6**	**93.3**
-contrastive learning	93.5	92.3	92.9
-filter	93.4	92.2	92.8
-filter-attention	93.4	92.0	92.7

V. CONCLUSION

In this paper, a novel approach is presented for relational triplet extraction, emphasizing a relation-first perspective. This work incorporates a contrastive strategy and a filter-attention mechanism to effectively address the challenges posed by redundant relations and the accurate classification of relations within a diverse range. The proposed model also enhances the synergy between subtasks and effectively filters out extraneous information during sequence tagging. Empirical evaluations conducted on publicly available datasets showcase the remarkable performance of CLFM. In the subsequent research endeavors, the exploration of more intricate filter-attention mechanism is on the horizon to elevate the overall quality of representations.

REFERENCES

[1] V. Gupta and G. S. Lehal, "A survey of text summarization extractive techniques," *Journal of emerging technologies in web intelligence*, vol. 2, no. 3, pp. 258–268, 2010.

[2] S. Riedel, L. Yao, A. McCallum, and B. M. Marlin, "Relation extraction with matrix factorization and universal schemas," *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: human language technologies*, pp. 74–84, 2013.

[3] D. Diefenbach, V. Lopez, K. Singh, and P. Maret, "Core techniques of question answering systems over knowledge bases: a survey," *Knowledge and Information systems*, vol. 55, pp. 529–569, 2018.

[4] D. Zelenko, C. Aone, and A. Richardella, "Kernel methods for relation extraction," *Journal of machine learning research*, vol. 3, no. Feb, pp. 1083–1106, 2003.

[5] Y. S. Chan and D. Roth, "Exploiting syntactico-semantic structures for relation extraction," *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pp. 551–560, 2011.

[6] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*, pp. 2335–2344, 2014.

[7] Z. Zhong and D. Chen, "A frustratingly easy approach for entity and relation extraction," *arXiv preprint arXiv:2010.12812*, 2020.

[8] Y. Yuan, X. Zhou, S. Pan, Q. Zhu, and L. Guo, "A relation-specific attention network for joint entity and relation extraction," *Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence IJCAI-PRICAI-20*, 2020.

[9] X. Zeng, D. Zeng, S. He, L. Kang, and J. Zhao, "Extracting relational facts by an end-to-end neural model with copy mechanism," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018.

[10] R. Panchendrarajan and A. Amaresan, "Bidirectional lstm-crf for named entity recognition," *The 32nd Pacific Asia Conference on Language, Information and Computation (PACLIC 32)*, 2019.

[11] P.-L. H. Cabot and R. Navigli, "Rebel: Relation extraction by end-to-end language generation," *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 2370–2381, 2021.

[12] H. Zheng, R. Wen, X. Chen, Y. Yang, Y. Zhang, Z. Zhang, N. Zhang, B. Qin, M. Xu, and Y. Zheng, "Prgc: Potential relation and global correspondence based joint relational triple extraction," *arXiv preprint arXiv:2106.09895*, 2021.

[13] L. Wu, J. Li, Y. Wang, Q. Meng, T. Qin, W. Chen, M. Zhang, T.-Y. Liu *et al.*, "R-drop: Regularized dropout for neural networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 10 890–10 905, 2021.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv*, 2017.

[15] S. Riedel, L. Yao, and A. K. Mccallum, "Modeling relations and their mentions without labeled text," *Springer-Verlag*, 2010.

[16] C. Gardent, A. Shimorina, S. Narayan, and L. Perez-Beltrachini, "Creating training corpora for nlg micro-planning," *Meeting of the Association for Computational Linguistics*, 2017.

[17] X. Yu and W. Lam, "Jointly identifying entities and extracting relations in encyclopedia text via a graphical model approach," *Coling 2010: Posters*, pp. 1399–1407, 2010.

[18] M. Miwa and Y. Sasaki, "Modeling joint entity and relation extraction with table representation," *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1858–1869, 2014.

[19] X. Ren, Z. Wu, W. He, M. Qu, C. R. Voss, H. Ji, T. F. Abdelzaher, and J. Han, "Cotype: Joint extraction of typed entities and relations with knowledge bases," *Proceedings of the 26th international conference on world wide web*, pp. 1015–1024, 2017.

[20] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, and B. Xu, "Joint extraction of entities and relations based on a novel tagging scheme," *arXiv preprint arXiv:1706.05075*, 2017.

[21] B. Yu, Z. Zhang, X. Shu, Y. Wang, T. Liu, B. Wang, and S. Li, "Joint extraction of entities and relations based on a novel decomposition strategy," *arXiv preprint arXiv:1909.04273*, 2019.

[22] Z. Wei, J. Su, Y. Wang, Y. Tian, and Y. Chang, "A novel cascade binary tagging framework for relational triple extraction," *arXiv preprint arXiv:1909.03227*, 2019.

[23] Y. Wang, B. Yu, Y. Zhang, T. Liu, H. Zhu, and L. Sun, "Tplinker: Single-stage joint extraction of entities and relations through token pair linking," *Proceedings of the 28th International Conference on Computational Linguistics*, 2020.

[24] Y. Shang, H. Huang, X. Sun, W. Wei, and X. Mao, "Relational triple extraction: One step is enough," *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pp. 4360–4366, 2022. [Online]. Available: https://doi.org/10.24963/ijcai.2022/605

[25] T. Nayak and H. T. Ng, "Effective modeling of encoder-decoder architecture for joint entity and relation extraction," in *AAAI*, 2020, pp. 8528–8535.

[26] L. Ma, H. Ren, and X. Zhang, "Effective cascade dual-decoder model for joint entity and relation extraction," *CoRR*, vol. abs/2106.14163, 2021. [Online]. Available: https://arxiv.org/abs/2106.14163

[27] Y. Shang, H. Huang, and X. Mao, "Onerel: Joint entity and relation extraction with one module in one step," *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pp. 11 285–11 293, 2022. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/21379

[28] K. Zhao, H. Xu, Y. Cheng, X. Li, and K. Gao, "Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction," *Knowledge-Based Systems*, vol. 219, p. 106888, 2021.

[29] J. Ning, Z. Yang, Y. Sun, Z. Wang, and H. Lin, "Od-rte: A one-stage object detection framework for relational triple extraction," *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11 120–11 135, 2023.

[30] D. Ye, Y. Lin, P. Li, and M. Sun, "Packed levitated marker for entity and relation extraction," *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pp. 4904–4917, 2022. [Online]. Available: https://doi.org/10.18653/v1/2022.acl-long.337

[31] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 4171–4186, 2019. [Online]. Available: https://doi.org/10.18653/v1/n19-1423

[32] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized BERT pretraining approach," *CoRR*, vol. abs/1907.11692, 2019. [Online]. Available: http://arxiv.org/abs/1907.11692

[33] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013.

[34] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computer Science*, 2014.

[35] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in adam," *CoRR*, vol. abs/1711.05101, 2017. [Online]. Available: http://arxiv.org/abs/1711.05101

[36] H. Ye, N. Zhang, S. Deng, M. Chen, C. Tan, F. Huang, and H. Chen, "Contrastive triple extraction with generative transformer," *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pp. 14 257–14 265, 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/17677

[37] X. Tian, L. Jing, L. He, and F. Liu, "Stereorel: Relational triple extraction from a stereoscopic perspective," *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 4851–4861, 2021.

Enhancing Airborne Disease Prediction: Integrating Deep Infomax and Self-Organizing Maps for Risk Factor Identification

Bhakti S. Pimpale, Dr. Anala A. Pandit
Department of Computer Application
Veermata Jijabai Technological Institute (V.J.T.I)
Mumbai, India

***Abstract*—Asthma poses a significant global public health concern, particularly in urban centers where environmental pollutants and variable weather patterns contribute to heightened prevalence and symptom exacerbation. The Deonar dumping ground, one of Mumbai's largest landfills, releases a complex mix of particulate matter and hazardous gases, posing a serious threat to local respiratory health. Despite the urgency for comprehensive research integrating patient-specific data with localized weather and air quality metrics, such studies remain limited. This study addresses the critical research gap by investigating asthma risk factors near the Deonar dumping ground. Integrating detailed patient records with precise local weather and air quality measurements, our research aims to unravel the intricate relationship between environmental exposure and respiratory health outcomes. The findings provide crucial insights into the specific risk factors influencing asthma incidence and severity in this region, informing the development of targeted interventions and mitigation strategies. Employing a novel ensemble Deep Info Max - Self-Organizing Map (DIM-SOM) technique, our study compares its performance with various clustering algorithms, including SOM, K-Means, Bisecting K-Means, DBSCAN, and others. The novel ensemble DIM-SOM demonstrated superior performance, achieving a significantly higher Silhouette Score of 0.9234, a lower Davies-Bouldin Score of 0.1276, and a more favorable Calinski-Harabasz Score of 389723.6225 compared to other algorithms. These findings underscore the efficacy of the novel ensemble DIM-SOM approach in generating dense, well-separated, and meaningful clusters, emphasizing its potential to enhance clustering performance compared to traditional algorithms. The study further emphasizes the need for proactive mitigation measures and tailored healthcare interventions based on the identified environmental risk factors.**

***Keywords*—Asthma; deepinfomax; self organizing map; risk factors; air pollution**

I. Introduction

Asthma, a chronic respiratory condition marked by airway inflammation and hyper-responsiveness, has been a focal point of global health concerns. Urban environments, characterized by a dynamic interplay of environmental pollutants and ever-changing weather patterns, witness a rising prevalence of asthma and an escalation of related symptoms [1] [2] [3] [4]. Mumbai, home to over 20 million residents, grapples with the intricate relationships between air pollution, meteorological factors, and public health. The Deonar dumping ground, among India's largest landfills, stands as a substantial contributor to air pollution, raising pertinent concerns about potential health risks, particularly respiratory ailments such as asthma.

In light of the advancements and challenges faced in this domain, this study seeks to contribute to the existing body of knowledge by investigating the risk factors associated with asthma in the vicinity of the Deonar dumping ground. Existing literature as discussed in Section II highlights the complexities of understanding the intricate relationship between environmental exposure and respiratory health outcomes. Despite a growing body of research, there remains a critical gap in studies that integrate patient-specific data with localized weather and air pollution metrics. This research endeavors to bridge this gap.

The importance of this study lies in its potential to offer crucial insights into the specific risk factors contributing to the incidence and severity of asthma in the specified region. By merging detailed patient records with precise measurements of local weather patterns and air quality indices, this research aims to unravel complex relationships, providing a nuanced understanding of the environmental determinants of respiratory health. The novelty of this approach is underscored by the proposed ensemble Deep Info Max - Self-Organizing Map (DIM-SOM) technique, offering a sophisticated methodology for asthma risk factor analysis.

While existing studies provide valuable information, this research distinguishes itself through its comprehensive integration of patient-specific data and localized environmental metrics, emphasizing the specificity of the Deonar dumping ground's impact on respiratory health. By shedding light on the complex interplay between environmental exposure and asthma outcomes, this study aims to provide actionable insights for local authorities and healthcare providers. Although such insights are essential for the development of targeted interventions, mitigation strategies, and informed healthcare protocols.

The subsequent sections of this paper provide a structured approach to the study. Section I introduces the significance of the study, emphasizing the global health concern posed by asthma, particularly in urban areas with high environmental pollutant levels. Section II offers a comprehensive overview of existing literature, summarizing related work in the field. Section III details the dataset used, incorporating patient-specific data, localized weather patterns, and air quality metrics. In Section IV, the proposed ensemble Deep Info Max - Self-Organizing Map (DIM-SOM) technique is introduced, outlining the methodology for asthma risk factor analysis. Section V describes the experimental setting, including environmental setup and machine setup. Following this, Section VI presents the results obtained from the clustering approach and initiates a comprehensive discussion on their implications. In the concluding Section VII, key findings are summarized, conclusions are drawn, and avenues for future research are suggested.

II. Related Work

Asthma's global prevalence and its association with urban environments, air pollution, and complex weather patterns have been extensively studied [5] [6] [7] [8]. The Deonar dumping ground's impact on the respiratory health of Mumbai's inhabitants highlights the critical need for an in-depth exploration of the specific risks involved in this context. Understanding the interplay of environmental factors and respiratory health outcomes in this setting is essential for the development of effective intervention strategies.

The study conducted in Helsinki [9] on the inhabitants of blockhouses built on a former dump area shed light on the potential health risks associated with landfill exposure. The findings suggested a slightly increased incidence of cancer, particularly in males following prolonged residence on the dump site. Moreover, the relative risk of chronic diseases, including asthma and chronic pancreatitis, exhibited a notable elevation. The implications of these results led to the subsequent demolition of the affected houses by the Helsinki City Council.

In a research report published by Shally Awasthi, Priya Tripathi, and Rajendra Prasad, [10] the authors aimed to identify asthma risk factors, including environmental exposures like motor vehicle air pollution, industrial smoke, active and passive smoking, and exposure to environmental tobacco smoke (ETS). The analysis, conducted according to Global Initiative for Asthma guidelines, categorized participants into two age groups, revealing significant associations between the studied environmental factors and asthma. The results indicated that motor vehicle air pollution, industrial smoke, and exposure to ETS were associated with asthma in participants aged 1–15 years. Additionally, the study highlighted the role of hospitalization in asthma severity.

Nonhlanhla Tlotleng, et al from Johannesburg [11] highlighted the health risks faced by waste recyclers, emphasizing the prevalence of acute respiratory symptoms in the population. Findings revealed that exposure to waste containing chemical residues significantly increased the likelihood of respiratory symptoms. Moreover, discrepancies in symptom prevalence were observed across different landfill sites, underscoring the need for improved occupational health and safety measures. Recommendations included providing appropriate protective gear and promoting hygiene practices to mitigate health hazards associated with waste sorting among informal workers.

ShriKant Singh, Praveen Chokhandre, Pradeep S. Salve, Rahul Rajak [12] aimed to evaluate the health effects of a dumping site on the nearby community, emphasizing potential risk factors associated with solid waste management. Utilizing a case comparison design, the research identified a notable increase in respiratory illnesses, eye irritation, and stomach problems among the exposed group in comparison to the non-exposed group. The findings underscore the significant impact of exposure to the dumping site on the prevalence of respiratory illness and eye infections, as highlighted by both the Propensity Score Matching (PSM) method and multivariate analysis. Furthermore, the assessment of air-quality-index indicated concerning levels of PM10 and PM2.5 during a fire outbreak at the Deonar dumping site, reinforcing the urgency for effective waste management strategies.

The study conducted by Seung-Woo Shin et al. [13] delves into the impact of air pollution levels on the severity of asthma exacerbations. Over a ten-year period, the research collected data from 143 adult asthmatics who experienced 618 exacerbation episodes, analyzing the influence of air pollutants such as O3, SO2, and NO2. The findings reveal significant associations between asthma exacerbations and increased pollutant levels during summer and winter, emphasizing similar relative risks for moderate and severe exacerbations. These results underscore the potential risks posed by specific air pollutants on asthma severity, particularly during distinct seasons, contributing valuable insights to the field of respiratory health. The study initiated by Kebalepile MM, Dzikiti LN, Voyi K. [14] investigated acute respiratory outcomes, particularly asthma, in relation to environmental exposures using self-organizing maps (SOM), a computational intelligence paradigm of artificial neural networks (ANNs). Utilizing air quality data such as nitrogen dioxide, sulphur dioxide, and particulate matter, along with clinical and socio-demographic information, the SOM effectively classified asthma outcomes. Notably, age emerged as a significant factor, with older patients exhibiting a higher likelihood of asthma. The study highlighted the importance of SO2 as a critical pollutant requiring attention. The SOM model demonstrated a low quantization error, suggesting its efficacy in studying asthma outcomes with multidimensional data. The overall accuracy of the model was found to be 59%.

In this study [15], the authors address the existing gap in mHealth applications for asthma self-management by proposing an optimized Deep Neural Network Regression (DNNR) model. Integrating weather, demography, and asthma tracking, the model demonstrates significant potential, achieving a score of 0.83 with Mean Absolute Error (MAE) of 1.44 and Mean Squared Error (MSE) of 3.62. The authors further enhance the model's accuracy through an optimization process, resulting in a remarkable 94% accuracy rate with MAE of 0.20 and MSE of 0.09.

Based on the literature review, it is evident that existing studies have explored the relationship between environmental factors and asthma prevalence and severity. However, there is a distinct gap in research that delves into personalized risk factor identification. While prior studies have provided valuable insights into the general associations between air pollution, climatic conditions, and asthma, this research seeks to address this gap by focusing on personalized risk, thus providing a more targeted approach to asthma management and prevention. This study's approach holds promise for identifying individualized risk factors and tailoring interventions for patients facing air pollution-induced asthma exacerbations.

III. Dataset

The data for this research was collected from multiple sources to facilitate a comprehensive analysis of the risk factors

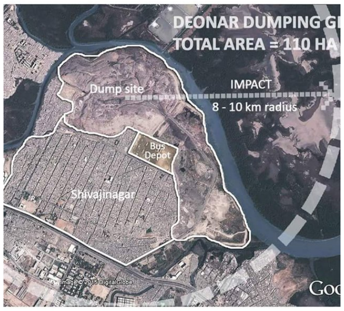


Fig. 1. Location of deonar dumping ground.

associated with asthma in the vicinity of Mumbai's Deonar dumping ground as shown in Fig. 1. The most affected areas near dumping ground are Bainganwadi, Deonar village, Govandi village, Shivajinagar, KamalaRamannagar, Rafiqnagar, Sanjaynagar and Shantinagar. Patients' data, air pollution and weather data is mainly collected from above selected areas.

A. Data Collection

1) Patient Data: Patient data, crucial for understanding the health profiles and medical histories of individuals, was obtained from Shatabdi Hospital, a government healthcare facility situated in Govandi, Mumbai. The patient data, comprising a comprehensive set of health metrics and demographic details, including Date of Visit, Age, Gender, Location, Smoking Status, Blood pressure, Body temperature, Height, Weight, Allergy status, and Symptoms of asthma (like cough intensity, sputum colour, wheezing, rales, rhonchus, etc.), spans the period from Jan 2015 to March 2020, encompassing a total of 46158 patients of asthma. The dataset contains total 76 features including physical characteristics and symptoms as discussed above. This dataset forms the basis for the study's examination of asthma incidence, symptomatology, and individual risk factors, enabling a detailed analysis of the interplay between air pollution and patient health.

2) Air Pollution Data: Air pollution data was sourced from the Copernicus Atmosphere Monitoring Service (CAMS) via the CAMS Global Reanalysis EAC4 dataset [16]. This dataset offers detailed information on air quality and atmospheric conditions. The collection period extends from Jan 2015 to March 2020, encompassing various pollutants and their concentrations, which are pivotal to assessing their impact on respiratory health in the study area. The dataset includes various key pollutants such as SO2, NO2, PM10, PM2.5, CO, O3, CH4, NH3, and dust aerosols, all of which were measured in kilograms per kilogram (kg/kg).

3) Weather Data: To understand the influence of meteorological factors on asthma risk, weather data was gathered from the NASA's Power Data Access Viewer [17]. The dataset provides a comprehensive overview of weather variables including temperature, humidity, Uv index, wind speed, dew point and rainfall etc. Data collection for weather variables aligns with the time-frame from Jan 2015 to March 2020, ensuring a holistic examination of weather patterns in the study region.

From Table I, it is clear that several parameters in the dataset display a positively skewed distribution, indicating a prevalence of higher values, while others exhibit a symmetrical skew, highlighting a more balanced distribution of values. The positively skewed parameters, including SO2, CO, NO2, O3, PM10, PM2.5, NH3, Dust Aerosol concentrations, UV index, wind speed, dew point, rainfall, and maximum temperature, exhibit a distribution where the tail of the data points extends towards higher values. This skewness indicates that, most of these parameters have relatively lower values, while a small number of data points have significantly higher values. Understanding the skewness of these parameters is crucial for assessing their impact on asthma exacerbation, as it suggests that elevated levels of these pollutants might be associated with more severe asthma symptoms. In contrast, parameters such as CH4 (methane), temperature, relative humidity, and minimum temperature demonstrate a more symmetrical distribution, where the data is relatively evenly distributed around the mean. This symmetric distribution implies that the values of these parameters do not have a strong skew towards higher or lower values. For these parameters, it is essential to explore their impact on asthma exacerbation by considering their overall levels rather than the skewness of their distribution.

B. Data Pre-processing

Prior to the analysis, a series of data pre-processing steps were applied to ensure the quality and integrity of the dataset.

1) Missing values: All the datasets were first examined for missing values. In the case of the air pollution dataset, missing values were handled using the KNN Imputer [18], which imputes missing values based on the values of the nearest neighbors in the feature space. This approach enabled to account for the complex interdependencies among the pollutant variables, ensuring a more accurate representation of the air quality indicators. Air pollution data contains only 73 missing values in UV index column shown in Fig. 2. A customized function was developed, to fill the missing values

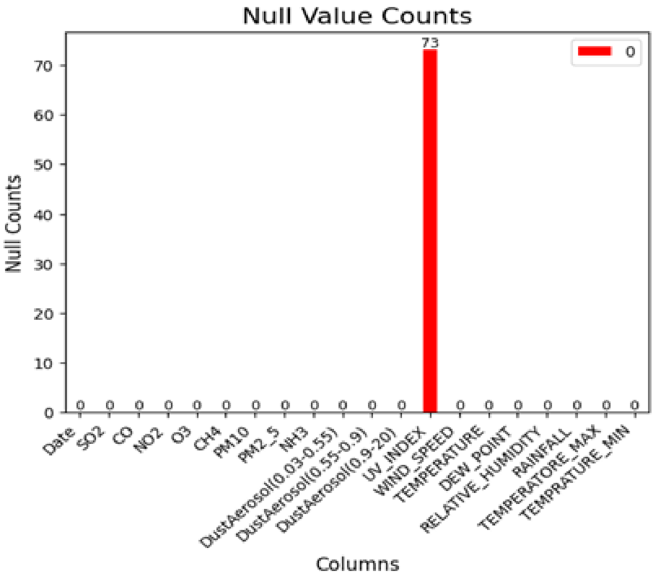


Fig. 2. Null value count of air pollution and weather data.

TABLE I. Statistical Representation of Air Pollution and Weather Data

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
SO2	2162.0	1.474753e-08	8.211345e-09	2.500000e-09	6.155000e-09	1.540000e-08	2.190000e-08	3.510000e-08
CO	2162.0	7.954561e-07	6.059048e-07	1.130000e-07	2.202500e-07	7.110000e-07	1.230000e-06	4.330000e-06
NO2	2162.0	1.247824e-08	6.983532e-09	2.100000e-09	5.760000e-09	1.200000e-08	1.800000e-08	3.680000e-08
O3	2162.0	8.128904e-08	4.077776e-08	2.770000e-08	4.110000e-08	7.870000e-08	1.130000e-07	2.560000e-07
CH4	2162.0	1.006156e-06	7.618521e-09	9.870000e-07	1.000000e-06	1.010000e-06	1.010000e-06	1.020000e-06
PM10	2162.0	2.339296e-07	1.675897e-07	3.510000e-08	8.315000e-08	1.775000e-07	3.670000e-07	8.600000e-07
PM2_5	2162.0	1.645160e-07	1.189400e-07	2.370000e-08	5.630000e-08	1.260000e-07	2.590000e-07	6.060000e-07
NH3	2162.0	4.738377e-10	3.942314e-10	3.530000e-11	1.310000e-10	4.320000e-10	6.490000e-10	2.300000e-09
DustAerosol(0.03-0.55)	2162.0	2.675921e-09	2.607848e-09	5.260000e-12	5.785000e-10	2.060000e-09	3.850000e-09	1.730000e-08
DustAerosol(0.55-0.9)	2162.0	5.110430e-09	5.022425e-09	6.930000e-12	9.210000e-10	3.790000e-09	7.790000e-09	3.210000e-08
DustAerosol(0.9-20)	2162.0	6.185520e-09	7.339973e-09	0.000000e+00	4.822500e-10	2.980000e-09	9.795000e-09	4.140000e-08
UV_INDEX	2162.0	1.752590e+00	5.417236e-01	2.100000e-01	1.330000e+00	1.700000e+00	2.200000e+00	3.100000e+00
WIND_SPEED	2162.0	2.509676e+00	1.006681e+00	8.600000e-01	1.830000e+00	2.230000e+00	2.900000e+00	8.270000e+00
TEMPERATURE	2162.0	2.677056e+01	2.874391e+00	1.725000e+01	2.507000e+01	2.638500e+01	2.881000e+01	3.436000e+01
DEW_POINT	2162.0	1.868569e+01	5.950174e+00	-2.540000e+00	1.438000e+01	1.998000e+01	2.410000e+01	2.590000e+01
RELATIVE_HUMIDITY	2162.0	6.707998e+01	1.847839e+01	1.681000e+01	5.320500e+01	6.565500e+01	8.619000e+01	9.525000e+01
RAINFALL	2162.0	7.148363e+00	1.765512e+01	0.000000e+00	0.000000e+00	5.000000e-02	4.400000e+00	1.438900e+02
TEMPERATORE_MAX	2162.0	3.289481e+01	4.051823e+00	2.487000e+01	2.948000e+01	3.208000e+01	3.619000e+01	4.491000e+01
TEMPRATURE_MIN	2162.0	2.216858e+01	3.766420e+00	1.051000e+01	1.912000e+01	2.339500e+01	2.502000e+01	2.891000e+01

of blood pressure and temperature in the patients' dataset. The function utilizes age-specific ranges for blood pressure and temperature, using linear interpolation to estimate missing values within the normal ranges. This approach ensures that the filled values remain within the expected physiological limits for each patient, minimizing potential data inaccuracies. Table II shows the missing value count in patients' dataset.

TABLE II. Missing Values in Patients' Data

Column	Null Count
Date	0
Diastolic Blood Pressure	3304
Systolic Blood Pressure	3320
Temperature	10141
.	.
.	.
Disease	0

2) Label encoding: To facilitate the integration of the Body Mass Index (BMI) information into the analysis, a categorical BMI variable was derived from the patients' body weight and height measurements. The BMI values were categorized into distinct groups, namely 'Underweight,' 'Normal,' and 'Overweight,' based on predefined threshold values. To incorporate this categorical information into the dataset, a one-hot encoding technique was applied. This process involved transforming the categorical variable into a set of binary variables, with each representing a specific BMI category.

3) Data normalization: To ensure consistent scaling in the dataset, an extensive analysis of various scaling methods was conducted. Initially, several known standardization techniques were explored, including the standard scaler and normalization, along with the application of the MinMaxScaler for assessing its efficacy in preserving data distribution within a specific range. However, after thorough experimentation and evaluation, the MaxAbsScaler [19] technique emerged as the most suitable choice. Results of all normalization techniques shown in Table V.

The MaxAbsScaler normalization method was selected for its remarkable ability to retain the data's inherent sparsity and preserve the relative relationships among the features. By rescaling each feature based on its maximum absolute value, the method ensured that the range of each feature fell within the [-1, 1] range. This approach not only maintained the dataset's unique characteristics but also enabled a fair comparison and interpretation of the features. Consequently, the adoption of the MaxAbsScaler technique not only provided a balanced and optimal scaling approach but also contributed to the robustness and reliability of subsequent analyses.

4) Data integration: The integration of the patient data and air pollution data involved a merging procedure based on the 'Date of Visit' and 'Date' variables from the respective datasets. This operation facilitated the alignment of the patients' symptomatology records with the corresponding air pollution data, providing insights into the potential associations between symptom onset and ambient air quality. A unified dataset was created, laying the groundwork for a detailed analysis of the interdependencies between air pollution exposure and asthma incidence.

5) Data correlation: The heatmap shown in Fig. 3 revealed several notable correlations within the dataset. Parameters such as runny nose, high cough, chest tightness, throat irritation, itchy eyes, and watery eyes exhibited a positive correlation with CO, O3, PM10, and PM2.5, indicative of potential associations between these respiratory symptoms and elevated levels of air pollutants. Conversely, these symptoms demonstrated a negative correlation with minimum temperature and Dew Point. The negative correlation indicate that cooler and more humid weather conditions could be linked to an increase in the severity of the mentioned respiratory issues.

IV. Proposed Model

In the initial stages of the study, the consolidated dataset, which included the merged information from air pollution, weather, and patient health records, underwent an extensive application of various clustering algorithms. K-means [20] [21], Self-Organizing Maps (SOM) [22], Mini-Batch K-means [23],FuzzyCmeans [24], Birch clustering [25], DBSCAN clustering [26], Agglomerative clustering [27], Spectral Clustering, Bisecting Kmeans clustering and Gaussian Mixture models [28] were each individually applied across lag 2 to 15, encompassing a comprehensive analysis of the dataset at different time intervals. Despite these efforts, the outcomes did not meet the predetermined benchmarks, indicating a need for an alternative approach. DeepInfomax algorithm [29], a deep learning

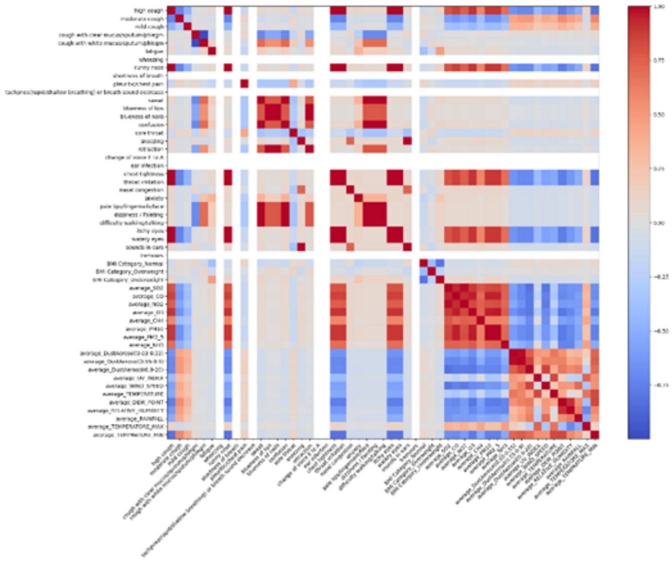


Fig. 3. HeatMap.

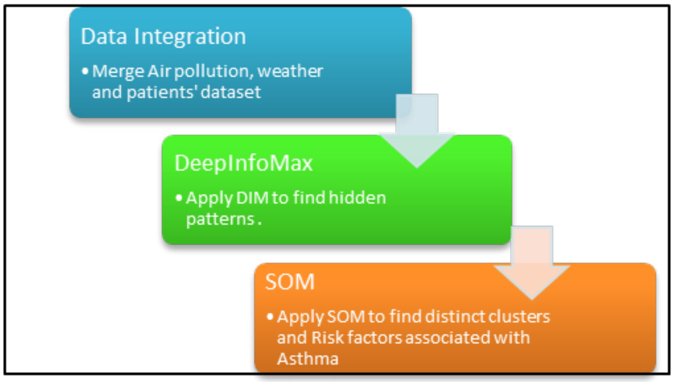

Data Integration
• Merge Air pollution, weather and patients' dataset
DeepInfoMax
• Apply DIM to find hidden patterns.
SOM
• Apply SOM to find distinct clusters and Risk factors associated with Asthma

Fig. 4. Proposed DIM-SOM Model.

technique was implemented. Although it was mainly used for images. This algorithm is capable of extracting intricate features and uncovering latent patterns within complex dataset. In the case of DIM, the model is trained to predict certain parts of the input data from other parts, without requiring explicit supervision from labeled data.

Using the power of the DeepInfomax technique, intricate associations and previously unnoticed patterns within the combined dataset numeric in nature were observed. This helps to better understand the complex relationships within the dataset, which traditional clustering techniques couldn't capture well.

Later the Self-Organizing Map (SOM) technique was used to further refine the clustering outcomes and establish distinct clusters and associated labels within the dataset. This combined methodology yielded a notable enhancement in the evaluation metric. This integrated approach not only improved the precision of the clustering process but also facilitated a holistic understanding of the multifaceted factors contributing to the incidence and patterns of asthma in relation to the interconnected influences of air pollution and weather conditions over the specified lag. Fig. 4 shows the proposed model with three different stages like data integration, pattern recognition and clustering followed by risk factor analysis.

V. Experimental Settings

The experiments were performed on an Intel Core i5 @ 1.19 GHz and 8 GB of memory. Python software was used for model creation and prototyping because it includes publicly accessible library sets for machine learning and statistical methods like Scikit-learn, and Matplotlib. Modeling tests were run to confirm the efficacy of the proposed model. Data and results were plotted using the Python 2D graphing tool included in the Matplotlib. The effectiveness of the model was examined using Sklearns cluster metrics, a Python module for performance measurements of machine learning models. All experiments were also conducted using Google Colab, a cloud-based Jupyter notebook environment that enabled seamless integration with Google Drive and provided access to high-performance computing resources. Validated our results using earlier method.

VI. Result and Discussion

The aim of this study was to identify clusters that could analyze the impact of air pollution on patients living near a dumping ground along with the associated risk factors. To accomplish this, determining the precise time-frame between disease onset and patient treatment was crucial. Multiple methods were utilized to determine the most efficient approach, and the results were documented and are presented in Table III and Table IV for further analysis.

However, the results obtained from these different methods did not reveal clear or meaningful patterns in the data. This unexpected outcome posed a challenge in selecting the appropriate lag for subsequent analysis. Therefore, identifying the best lag remains a crucial aspect of the research, prompting to explore alternative approaches to address this issue.

To determine the optimal lag, the novel technique of creating an ensemble using the Deep Info Max (DIM) initially, followed by the application of the self-organizing map (SOM) was implemented. The choice of the SOM algorithm was based on its ability to maintain the topological characteristics of the data, thereby uncovering the fundamental structure and interrelationships among data points. This capability becomes particularly pertinent when dealing with intricate, high-dimensional datasets, where preserving the original data structure is paramount. Unlike the K-Means algorithm, SOM excels in capturing non-linear associations within the data, proving invaluable for datasets where clusters may lack clear boundaries or exhibit complex configurations. Additionally, SOM demonstrates greater resilience to outliers and noise, thereby minimizing the influence of such irregularities on the clustering outcomes. Moreover, SOM can function as a dimensionality reduction technique during the clustering process, which proves especially advantageous when dealing with datasets featuring numerous dimensions(in this case 97 features). This reduction can simplify data exploration and comprehension, facilitating a more accessible understanding of the data's underlying patterns and relationships. Architecture of DIM is shown in Fig. 5 The DIM neural network architecture comprises an input layer designed to accommodate data

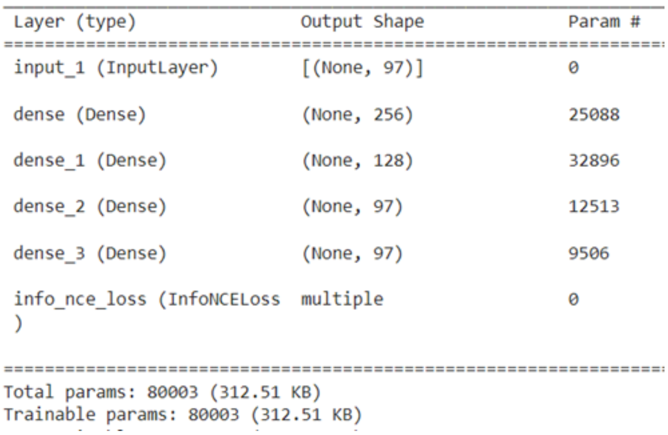
TABLE III. Silhouette Scores for Kmeans, Minibatch kmeans, SOM, Gaussian Mixture and FuzzyCMeans

Silhouette Index					
Lag	Kmeans	Minibatch Kmeans	Self-organizing maps	Gaussian Mixture	FuzzyCmeans
2	0.199	0.272	0.260	0.181	0.242
3	0.270	0.196	0.259	0.180	0.242
4	0.270	0.203	0.260	0.121	0.242
5	0.271	0.193	0.260	0.186	0.244
6	0.272	0.186	0.259	0.184	0.245
7	0.272	0.203	0.259	0.184	0.245
8	0.274	0.269	0.259	0.178	0.246
9	0.276	0.199	0.259	0.176	0.248
10	0.278	0.204	0.260	0.177	0.249
11	0.279	0.237	0.261	0.176	0.250
12	0.280	0.203	0.262	0.152	0.250
13	0.281	0.207	0.262	0.192	0.251
14	0.281	0.200	0.264	0.179	0.251
15	0.282	0.211	0.265	0.190	0.252

TABLE IV. Silhouette Scores for Birch, DBSCAN, Agglomerative, Spectral and Bisecting Kmeans Clustering

Silhouette Index					
Lag	Birch	DBSCAN	Agglomerative	Spectral	Bisecting kmeans
2	0.153	-0.0935	0.317	0.318	0.298
3	0.157	-0.1047	0.319	0.321	0.301
4	0.159	-0.0914	0.321	0.320	0.301
5	0.162	-0.1105	0.321	0.321	0.301
6	0.163	-0.0999	0.322	0.324	0.300
7	0.163	-0.0999	0.322	0.324	0.300
8	0.169	-0.1067	0.327	0.325	0.298
9	0.171	-0.0987	0.327	0.328	0.297
10	0.173	-0.1031	0.329	0.329	0.302
11	0.218	-0.0981	0.330	0.328	0.308
12	0.219	-0.1179	0.327	0.330	0.310
13	0.220	-0.1145	0.333	0.330	0.310
14	0.219	-0.0992	0.331	0.331	0.312
15	0.218	-0.0976	0.333	0.331	0.312

instances of 97 features. This is followed by a sequence of four densely connected layers, namely, ‘dense’, ‘dense1’, ‘dense2’, and ‘dense3’, responsible for hierarchically processing the input data. These layers utilize various mathematical operations, including weighted summation and activation functions, to extract and transform the input data into meaningful representations. Furthermore, the inclusion of a custom ‘InfoNCELoss’ layer underscores the network’s reliance on the Information Noise-Contrastive Estimation (InfoNCE) loss function. By leveraging this tailored layer, the model is adept at maximizing the agreement between representations of related instances while minimizing the agreement between representations of unrelated instances. This approach, commonly associated with self-supervised learning, facilitates the acquisition of effective data representations, consequently enabling the model to find intricate patterns and relationships within the dataset.



```
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 97)]              0

 dense (Dense)               (None, 256)               25088

 dense_1 (Dense)             (None, 128)               32896

 dense_2 (Dense)             (None, 97)                12513

 dense_3 (Dense)             (None, 97)                9506

 info_nce_loss (InfoNCELoss  multiple                  0
 )

=================================================================
Total params: 80003 (312.51 KB)
Trainable params: 80003 (312.51 KB)
```

Fig. 5. Architecture of DIM.

Let $\mathbf{X}$ denote the input data with dimensions $(N, 97)$, where N represents the batch size. The neural network architecture can be symbolically represented as follows:

$$\text{Input Layer:} \quad \mathbf{X} \in \mathbf{R}^{N\times 97} \tag{1}$$

$$\text{Dense Layer 1:} \quad \mathbf{H}^{(1)} = \sigma(\mathbf{X}\cdot\mathbf{W}^{(1)} + \mathbf{b}^{(1)}) \tag{2}$$

$$\text{Dense Layer 2:} \quad \mathbf{H}^{(2)} = \sigma(\mathbf{H}^{(1)}\cdot\mathbf{W}^{(2)} + \mathbf{b}^{(2)}) \tag{3}$$

$$\text{Dense Layer 3:} \quad \mathbf{H}^{(3)} = \sigma(\mathbf{H}^{(2)}\cdot\mathbf{W}^{(3)} + \mathbf{b}^{(3)}) \tag{4}$$

$$\text{Dense Layer 4:} \quad \mathbf{H}^{(4)} = \sigma(\mathbf{H}^{(3)}\cdot\mathbf{W}^{(4)} + \mathbf{b}^{(4)}) \tag{5}$$

$$\text{InfoNCELoss Layer:} \quad \text{Custom layer utilizing InfoNCE} \tag{6}$$

In the above representation, σ denotes the activation function, such as the LeakyReLU function, which introduces non-linearities into the network. $W^{(i)}$ represents the weight matrix and $b^{(i)}$ denotes the bias vector for the ith dense layer.

The results of the novel ensemble Deep Info Max - Self-Organizing Map (DIM-SOM) technique, presented in Table VI, reflect the evaluation based on prominent clustering metrics, including the Silhouette Score [30], Calinski Harabasz Score [31], and Davies Bouldin Score [32]. The findings illustrate the complex dynamics involved in choosing the ideal lag value.

Among the considered lag values, the analysis spotlights lag 12 as a prominent contender with an impressive Silhouette Score of 0.9234, indicating distinct and well-defined clusters within the dataset. Moreover, the notably high Calinski-Harabasz Score of 389723.6225 for lag 12 further signifies the presence of dense and well-separated clusters, in producing meaningful clustering outcomes. Complementing these findings, the relatively low Davies-Bouldin Score of 0.1276 for lag 12 underlines the improved cluster separation compared to other lag values. Hence novel ensemble DIM-SOM technique, emphasizing its potential to generate compact, well-separated, and distinct clusters was established.

Further insight into the clustering performance was gained through a comparison of the self-organizing map (SOM) and the novel ensemble DIM-SOM. Table VII illustrates the comparison, highlighting the superior performance of the novel ensemble DIM-SOM over the standard SOM algorithm. Specifically, the novel ensemble DIM-SOM algorithm attained a significantly higher Silhouette Score of 0.9234, a lower Davies-Bouldin Score of 0.1276, and a more favorable Calinski-Harabasz Score of 389723.6225, surpassing the respective metrics achieved by the SOM algorithm (Silhouette Score: 0.2619, Davies-Bouldin Score: 1.6610, Calinski-Harabasz Score: 3040.4192). These findings underscore the efficacy of the novel ensemble DIM-SOM approach in generating dense, well-separated, and meaningful clusters, further emphasizing its potential to enhance clustering performance in comparison to the traditional SOM algorithm.

TABLE V. Result of Different Normalization Methods using DIM-SOM

Novel ensemble DIM - SOM using MinMaxScaler			
Lag	**Silhouette Score**	**Calinski-Harabasz Score**	**Davies-Bouldin Score**
2	0.64	14955.70	0.64
3	0.47	13708.11	0.76
4	0.47	13708.11	0.76
5	0.76	21635.34	0.44
6	0.57	19263.01	0.59
7	0.88	151696.15	0.17
8	0.45	8659.86	1.01
9	0.72	44186.74	0.39
10	0.62	9825.42	0.56
11	0.59	19719.31	0.64
12	0.34	4541.49	1.36
13	0.44	7187.50	0.99
14	0.49	8625.06	0.98
15	0.57	21352.39	0.62
Novel ensemble DIM - SOM using Standard Scaler			
Lag	**Silhouette Score**	**Calinski-Harabasz Score**	**Davies-Bouldin Score**
2	0.64	24194.60	0.49
3	0.60	15814.63	0.60
4	0.57	18366.48	0.55
5	Nan	Nan	Nan
6	Nan	Nan	Nan
7	0.56	17819.22	0.57
8	Nan	Nan	Nan
9	0.57	13829.26	0.62
10	Nan	Nan	Nan
11	Nan	Nan	Nan
12	Nan	Nan	Nan
13	Nan	Nan	Nan
14	Nan	Nan	Nan
15	Nan	Nan	Nan
Novel ensemble DIM - SOM using Normalizer(L1)			
Lag	**Silhouette Score**	**Calinski-Harabasz Score**	**Davies-Bouldin Score**
2	0.97	806.28	1.07
3	0.45	905.65	1.04
4	0.35	750.41	1.09
5	0.94	1466.52	1.09
6	0.70	1000.87	1.05
7	-0.06	865.39	1.06
8	0.56	895.44	1.09
9	0.72	799.64	1.06
10	-0.06	937.34	1.05
11	0.10	996.51	1.05
12	0.68	791.87	1.05
13	**1.00**	**1017.05**	**1.06**
14	0.06	920.00	1.07
15	0.40	1127.97	1.04
Novel ensemble DIM - SOM using Normalizer(L2)			
Lag	**Silhouette Score**	**Calinski-Harabasz Score**	**Davies-Bouldin Score**
2	-0.06	1806.77	1.03
3	-0.14	956.99	1.08
4	0.41	650.60	1.08
5	Nan	Nan	Nan
6	Nan	Nan	Nan
7	Nan	Nan	Nan
8	-0.06	1132.53	1.05
9	0	1227.77	1.08
10	**1**	**754.18**	**1.03**
11	0.29	649.79	1.04
12	0.37	852.65	1.03
13	-0.26	558.91	1.04
14	Nan	Nan	Nan
15	0.61	16888.15	0.53
Novel ensemble DIM - SOM using Normalizer(Max)			
Lag	**Silhouette Score**	**Calinski-Harabasz Score**	**Davies-Bouldin Score**
2	Nan	Nan	Nan
3	0.57	13159.36	0.63
4	0.56	1148.30	1.04
5	-0.05	920.83	1.05
6	0.00	682.11	1.02
7	0.60	17444.69	0.54
8	0.64	17009.87	0.55
9	Nan	Nan	Nan
10	**1.00**	**932.70**	**1.09**
11	0.92	875.73	1.04
12	0.62	19727.16	0.52
13	0.00	1040.48	1.12
14	Nan	Nan	Nan
15	Nan	Nan	Nan

TABLE VI. Novel Ensemble DIM - SOM using MaxAbsScaler

Novel ensemble DIM - SOM using MaxAbsScaler			
Lag	**Silhouette Score**	**Calinski-Harabasz Score**	**Davies-Bouldin Score**
2	0.7441	58304.0288	0.3750
3	0.8456	87980.6821	0.2492
4	0.3255	3466.9936	1.5385
5	0.5793	20339.4434	0.5563
6	Nan	Nan	Nan
7	Nan	Nan	Nan
8	0.4056	8932.7228	0.9797
9	0.6715	33287.8864	0.4172
10	Nan	Nan	Nan
11	Nan	Nan	Nan
12	**0.9234**	**389723.6225**	**0.1276**
13	0.4699	12359.7120	0.7685
14	0.5233	15293.3356	0.6500
15	0.6336	19533.1530	0.6197

TABLE VII. Comparison of Novel Ensemble DIM-SOM and SOM

Comparison of Novel Ensemble DIM-SOM and SOM			
Model	Silhouette Score	Calinski-Harabasz Score	Davies-Bouldin Score
DIM-SOM	**0.9234**	**389723.6225**	**0.1276**
SOM	0.2619	3040.4192	1.6610

Risk factor analysis The role of environmental factors in exacerbating asthma symptoms remains a critical area of investigation. The analysis focused on exploring the influence of air pollution, meteorological conditions, and obesity on two distinct clusters of asthma patients. The feature importance analysis is shown in Fig. 7 which notably indicate the varying impact of these factors on the identified clusters, describing important insights into the complex interactions between health outcomes and environmental conditions. The visualization from Fig. 6 clearly indicates the division of the data into two distinct clusters. Subsequent analysis enabled the classification of these clusters as “Asthma Aggravated by Air Pollution” and “Asthma Not Significantly Affected by Air Pollution”.

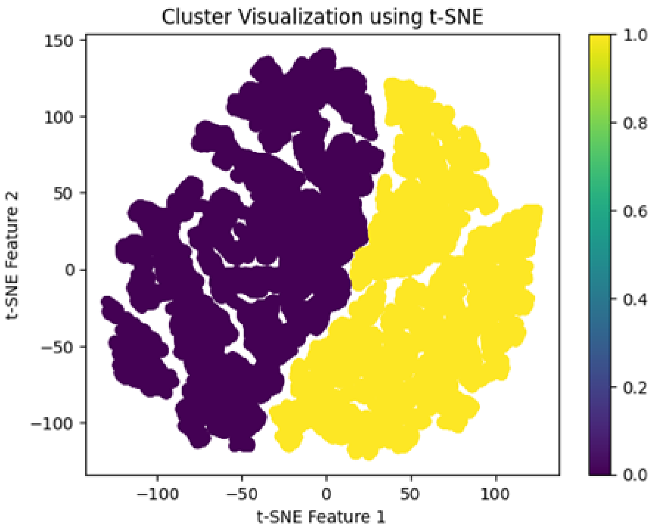


Fig. 6. Cluster visualization.

Asthma Aggravated by Air Pollution Cluster: The analysis revealed a significant association between air pollution and the severity of asthma symptoms in this cluster. The feature importance analysis highlighted the substantial impact of various pollutants, including average SO2, average CO, average NO2, average O3, average CH4, average PM10, and average PM2.5. Additionally, meteorological factors such as average TEMPERATURE, average DEW POINT, and average RELATIVE HUMIDITY exhibited a relationship with air pollutants, amplifying the adverse effects on asthma exacerbation. Asthma Not Significantly Affected by Air Pollution Cluster: In this cluster, the influence of air pollution on asthma symptoms appeared notably milder compared to the first cluster. Nevertheless, specific pollutants, particularly average PM10 and average PM2.5, still demonstrated a discernible impact, albeit to a very lesser extent. Moreover, meteorological parameters, including average TEMPERATURE, average WIND SPEED, and average RELATIVE HUMIDITY, played a prominent role, indicating their potential contribution to the manifestation of asthma symptoms. Obesity has long been recognized as a risk factor for various health conditions, including asthma. The feature importance analysis in both clusters elucidated the noteworthy impact of different BMI categories, emphasizing the crucial role of weight status in the severity and exacerbation of asthma symptoms. The analysis revealed the following insights: In both clusters, the feature importance values for both BMI Category Overweight and BMI Category Underweight suggested a similar negative impact, indicating that being either underweight or overweight may contribute to the aggravation of asthma symptoms (Table VIII).

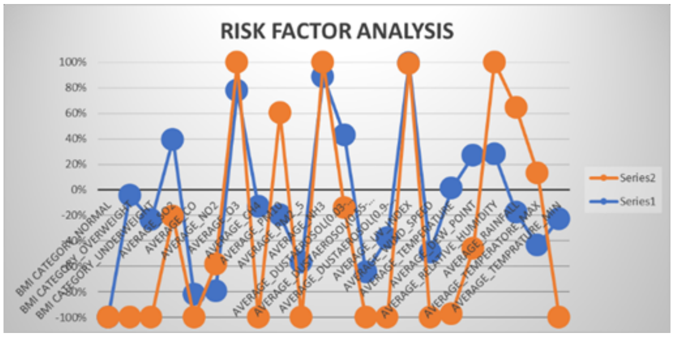


Fig. 7. Risk factor analysis.

TABLE VIII. COMPARISON OF STUDIES AND PROPOSED MODEL

Study	Focus	Key Findings	Results
Helsinki[9]	Landfill Exposure	Increased cancer incidence; elevated relative risk of chronic diseases, including asthma.	High correlation of air pollutants and pulmonary diseases.
Awasthi et al.[10]	Asthma Risk Factors	Significant associations between motor vehicle air pollution, industrial smoke, and asthma in participants aged 1–15 years.	Children are more vulnerable to the adverse effects of high air pollution.
Tlotleng et al.[11]	Waste Recyclers' Health	Increased respiratory symptoms in waste recyclers due to exposure to waste containing chemical residues.	Noticable correlation of respiratory symptoms and air pollution in the vicinity of dumping ground.
Singh et al.[12]	Dumping Site Impact	Notable increase in respiratory illnesses, eye irritation, and stomach problems among the exposed group as compared to controlled group.	High correlation of symptoms and air pollution.
Shin et al.[13]	Air Pollution and Asthma	Significant associations between asthma exacerbations and increased pollutant levels during summer and winter.	High correlation of air pollutants and weather parameters with asthma symptoms.
Kebalepile et al.[14]	SOM for Asthma Outcomes	SOM classified asthma outcomes based on air quality data and socio-demographic information. Age was a significant factor.	Classification Accuracy: 59%
Model for regression[15]	mHealth Application	Optimized DNNR model for asthma self-management.	Achieved 94% accuracy.
Proposed Model DIM-SOM	**Clustering Performance,for Asthma risk factor analysis**	**Higher Silhouette Score, lower Davies-Bouldin Score, and favorable Calinski-Harabasz Score for DIM-SOM.**	**Silhouette Score:0.92, Davies-Bouldin Score:0.12, Calinski-Harabasz Score:389723.62**

VII. CONCLUSION

This study has undertaken a comprehensive exploration of the intricate interplay between environmental factors and their impact on respiratory health, particularly in the context of the Deonar dumping ground in Mumbai. By employing a sophisticated novel ensemble model, which combines patient-specific data, localized weather patterns, air quality metrics, and the innovative ensemble Deep Info Max - Self-Organizing Map (DIM-SOM) technique, the multifaceted risk factors contributing to the incidence of asthma within this region were identified. The findings have unveiled the pivotal role played by environmental pollutants and meteorological variations in exacerbating(synonym) respiratory ailments, highlighting the heightened vulnerability of the local population to asthma. The novel ensemble DIM-SOM algorithm, introduced in this literature has demonstrated its efficacy in generating meaningful clusters, signifying its potential for accurate and interpretable insights into the complex relationship between environmental exposure and respiratory health. In the wake of these insights, this research calls for the implementation of proactive mitigation measures and tailored healthcare interventions to address the specific challenges posed by the Deonar dumping ground. By incorporating the results into policy planning and public health initiatives, local authorities and healthcare stakeholders can design effective strategies to alleviate the impact of environmental hazards on respiratory well-being in the affected communities.

Moreover, this study offers the potential for developing sustainable, evidence-based interventions that can safeguard public health in regions confronted with similar environmental challenges, further reinforcing the importance of ensemble methodologies in complex data analysis and pattern recognition within the context of public health.

While this study provides crucial insights into the complex interplay between environmental factors and respiratory health in the Deonar dumping ground area, it is important to acknowledge certain limitations. The reliance on retrospective data poses constraints on the establishment of causal relationships, warranting further longitudinal investigations to ascertain the temporality and directionality of the identified associations. Additionally, the study's focus on a specific geographical location necessitates caution in generalizing the findings to broader populations, emphasizing the need for multi-site studies to enhance the external validity of the results.

In terms of future scope, the incorporation of genetic and epigenetic data in conjunction with environmental factors could offer a more nuanced understanding of the individual susceptibility to asthma in polluted environments. Moreover, creating specific plans to help the community get involved and actively participate could be very effective in dealing with the many challenges caused by the environmental issues and respiratory health in similar areas.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Dr. Priyanka Kulkarni BAMS, MD - Ayurveda, and, Dr. Sonali Renushe BAMS, MD - Ayurveda for their essential guidance and support in assisting us to comprehend the medical component of our research work. Our research has been substantially enhanced by their insights into it.

REFERENCES

[1] Y. Zhang, X. Yin, and X. Zheng, *"The relationship between PM2.5 and the onset and exacerbation of childhood asthma: a short communication," Frontiers in Pediatrics,Frontiers Media SA,* vol. 11. August, 2023.

[2] M. Barnthouse, BL. Jones,*"The impact of environmental chronic and toxic stress on asthma," Clin Rev Allergy Immunol.* vol.57,no.3, pp.427–38,2019.

[3] Lara Joanna Macedo Borges, *"The Effect Of Biomedical Waste Disposal In Surrounding Areas Of Deonar Dumping Ground," International Journal Of Legal Developments And Allied Issues*, Vol. 8, no. 1,pp. 128-165, 2022.

[4] P. Tripathy and C. McFarlane, *"Perceptions of atmosphere: Air, waste, and narratives of life and work in Mumbai"*, *Environment and Planning D: Society and Space*, vol. 40, no. 4, pp. 664-682, 2022.

[5] A. Abbah, S. Xu, and A. Johannessen, *"Long-term exposure to outdoor air pollution and asthma in low- and middle-income countries: A systematic review protocol," PLoS One*, vol. 18, no. 7, 2023.

[6] D. Singh, I. Gupta, and A. Roy, *"The association of asthma and air pollution: Evidence from India"*, *Economics and Human Biology*, vol. 51, p. 101278, 2023.

[7] C. Hoffmann, M. Maglakelidze, E. von Schneidemesser et al., *"Asthma and COPD exacerbation in relation to outdoor air pollution in the metropolitan area of Berlin, Germany"*, *Respir Res*, vol. 23, p. 64, 2022.

[8] J. Chatkin, L. Correa, U. Santos, *"External Environmental Pollution as a Risk Factor for Asthma"*, *Clinical Reviews in Allergy & Immunology*, vol. 62, pp. 1-18, 2022.

[9] E. Pukkala, A. Pönkä, *"Increased incidence of cancer and asthma in houses built on a former dump area"*, *Environ Health Perspect*, vol. 109, no. 11, pp. 1121-1125, 2001.

[10] A. Lotfata, M. Moosazadeh, M. Helbich, and B. Hoseini,*"Socioeconomic and environmental determinants of asthma prevalence: a cross-sectional study at the U.S. County level using geographically weighted random forests."*, *International Journal of Health Geographics*, Springer Science and Business Media LLC, vol. 22, no. 1, 2023.

[11] N. Tlotleng et al., *"Prevalence of Respiratory Health Symptoms among Landfill Waste Recyclers in the City of Johannesburg, South Africa."*, *International journal of environmental research and public health*, vol. 16, no. 21, p. 4277, 2019.

[12] S. Singh, P. Chokhandre, P. Salve, and R. Rajak *"Open dumping site and health risks to proximate communities in Mumbai, India: A cross-sectional case-comparison study."*, *Clinical Epidemiology and Global Health*, vol. 9, 21, pp. 34-40, 2020.

[13] S. Shin et al., *"Effects of air pollution on moderate and severe asthma exacerbations."*, *The Journal of asthma : official journal of the Association for the Care of Asthma*, vol. 57, no. 8, pp. 875-885, 2020.

[14] M. Kebalepile, L.Dzikiti, and K. Voyi, *"Supervised Kohonen Self-Organizing Maps of Acute Asthma from Air Pollution Exposure."*, *International Journal of Environmental Research and Public Health*, MDPI AG, vol. 18, no. 21. p. 11071,2021.

[15] Haque R et al., *"Optimised deep neural network model to predict asthma exacerbation based on personalised weather triggers."*, *F1000Res*, vol. 10, pp. 911, 2021.

[16] CAMS. Available online: https://atmosphere.copernicus.eu/data (accessed on 28 August 2023).

[17] NASA. Available online: https://power.larc.nasa.gov/data-access-viewer/ (accessed on 28 August 2023).

[18] Murti, D. Prawidya, U. Pujianto, A. Wibawa, and M. Akbar. *"K-Nearest Neighbor (K-NN) based Missing Data Imputation," 2019 5th International Conference on Science in Information Technology (ICSITech)*,pp. 83-88.

[19] K. N. Abd Halim, A. S. Mohd Jaya, and A. F. A. Fadzil, *"Data Pre-Processing Algorithm for Neural Network Binary Classification Model in Bank Tele-Marketing," International Journal of Innovative Technology and Exploring Engineering,* vol. 9, no. 3, pp. 272–277, Jan. 30, 2020.

[20] S. P. Lloyd, *Least squares quantization in PCM.* Technical Report RR-5497, Bell Lab, September 1957.

[21] J. B. MacQueen, *"Some methods for classification and analysis of multivariate observations".* In L. M. Le Cam & J. Neyman (Eds.), Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, California: University of California Press,Vol. 1, pp. 281–297 , 1967.

[22] T. Kohonen, "The self-organizing map," *in Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464-1480, Sept. 1990.

[23] D. Sculley, *"Web-scale k-means clustering," Proceedings of the 19th international conference on World wide web.* ACM, Apr. 26, 2010.

[24] J. Bezdek *"Fuzzy C-means cluster analysis", Scholarpedia,* vol. 6, no. 7, pp. 2057, 2011.

[25] T. Zhang, R. Ramakrishnan, and M. Livny,*"BIRCH, " Proceedings of the 1996 ACM SIGMOD international conference on Management of data - SIGMOD '96.* ACM Press, 1996.

[26] M. Ester, H. Kriegel, J. Sander, and Xiaowei Xu, *"A density-based algorithm for discovering clusters in large spatial databases with noise," In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96).* AAAI Press, pp.226–231.1996.

[27] F. Murtagh and P. Legendre, *"Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?," Journal of Classification,* Springer Science and Business Media LLC Oct. vol. 31, no. 3, pp. 274–295, 2014.

[28] D. Reynolds, *"Gaussian Mixture Models," Encyclopedia of Biometrics.* Springer US, pp. 659–663, 2009.

[29] R. D. Hjelm et al., *"Learning deep representations by mutual information estimation and maximization." arXiv*, pp. 1-24, 2018.

[30] P. J. Rousseeuw,*"Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics*,Elsevier BV, vol. 20 , pp. 53–65, Nov. 1987.

[31] T. Calinski and J. Harabasz,*"A dendrite method for cluster analysis," Communications in Statistics - Theory and Methods*, vol. 3, no. 1. Informa UK Limited, pp. 1–27, 1974.

[32] D. L. Davies and D. W. Bouldin, *"A Cluster Separation Measure," IEEE Transactions on Pattern Analysis and Machine Intelligence*,Institute of Electrical and Electronics Engineers (IEEE), vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.

Enhancing Assamese Word Recognition for CBIR: A Comparative Study of Ensemble Methods and Feature Extraction Techniques

Naiwrita Borah*[1], Udayan Baruah[2], Barnali Dey[3], Merin Thomas[4], Sunanda Das[5], Moumi Pandit[6],
Bijoyeta Roy[7], Amrita Biswas[8]
PhD Scholar, Department of IT, SMIT, SMU, Sikkim, India[1];
Assistant Professor, Department of CSE, Presidency University, Bangalore, India[1]
Academic Registrar (in-charge) and Controller of Examinations,
Birangana Sati Sadhani Rajyik Vishwavidyalaya (A Government of Assam University), Golaghat, Assam, India[2]
Assistant professor (SG), Department of IT, SMIT, SMU, Sikkim, India[3]
Associate Professor, School of CSE, RV University, Bengaluru, India[4]
Department of CSE, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, India[5]
Associate professor, Department of EEE, SMIT, SMU, Sikkim, India[6]
Assistant professor, Department of CSE, SMIT, SMU, Sikkim, India[7]
Department of CSE, SMIT, SMU, Sikkim, India[8]

***Abstract*—This study conducts a thorough assessment of ensemble machine learning methods, specifically focusing on the identification of Assamese words. This task is crucial for improving Content-Based Image Retrieval systems and safeguarding the digital heritage of Assamese culture. We analyze the efficacy of different algorithms, such as CatBoost, XGBoost, Gradient Boosting, Random Forest, Bagging, AdaBoost, Stacking, and Histogram-Based Gradient Boosting, by thoroughly examining their performance in terms of accuracy, precision, recall, Kappa, F1-score, Matthews Correlation Coefficient, and AUC. The CatBoost algorithm stands out as the top performer, achieving an accuracy rate of 97.7%, precision rate of 95%, and recall rate of 96%. XGBoost is also acknowledged for its substantial effectiveness. This comparative analysis emphasizes CatBoost's superiority in terms of precision and recall. Additionally, it underscores the strong ability of ensemble classifiers to enhance assistive technologies, promote social inclusivity, and seamlessly integrate the Assamese language into technological applications.**

***Keywords*—Assamese literary works; automatic word recognition; comparative analysis; feature-based approaches; intelligent assistive technology; machine learning; word image analysis**

I. Introduction

Assamese, predominantly spoken in the state of Assam and various other regions in Northeast India, holds a significant position as one of the primary languages in India [1]. Hence, the development of a precise Assamese automatic word recognition system holds the potential to safeguard the cultural heritage of India. The recognition of Assamese words is of utmost importance in the preservation and promotion of the Assamese language. An accurate recognition system has the potential to address numerous domains, including digital resource management, the creation of educational tools, and the preservation of digital languages [2]. The ability to accurately recognize words enhances the accessibility of information for individuals who speak Assamese. The utilization of this technology enables the advancement of various technological applications, including content based image retrieval (CBIR), natural language processing (NLP), and information retrieval systems that are specifically tailored for the Assamese language. This facilitates the utilization of digital content, retrieval of pertinent information, and engagement in online platforms by Assamese speakers in their mother tongue.

The Assamese language exhibits dialectal variations, accents, and regional distinctions, which may pose difficulties in word recognition. Ensemble techniques refer to the automated process of effectively identifying and accurately interpreting word images [3]. This is achieved by integrating multiple models or classifiers, thereby leveraging their respective strengths and weaknesses. Ensemble methods have the ability to utilize the combined knowledge and expertise of individual models in order to generate results that are more accurate [4]. The task at hand pertains to the development of computational models, algorithms, and systems with the capability to effectively identify and comprehend Assamese words. In contrast to state-of-the-art methods, ensemble methods exhibit greater accuracy and robustness due to their utilization of multiple models trained on distinct subsets of the data or employing diverse feature representations. In order to facilitate the preservation and expansion of the Assamese language, as well as its integration into society, numerous researchers have employed various methodologies.

A. Distinctive Contribution

This study encompasses the completion of three primary objectives:

1) Identification of handcrafted features for Assamese Word Recognition.
2) Implementation of Ensemble Classification using the handcrafted features.

*Corresponding authors

3) Evaluation of performance to determine the optimal ensemble classification method for classification analysis.

The schematic depicted in Fig. 1 offers a graphical illustration of the sequence of tasks involved in this work.

This work is organised in the following format . Section II entails a thorough examination of relevant studies to establish the essential context and background for the research. Section III, entitled “Materials and Methods”, offers a thorough elucidation of the process by which the dataset was generated and the precise methodologies employed for feature engineering. Section IV, of the research paper discusses various ensemble models, while Section V, provides comprehensive details on the performance metrics employed to assess these models. Section VI, presents a comprehensive analysis of the findings, offering a comparative assessment of the effectiveness of the models. In Section VII, the paper concludes by presenting a succinct overview of the primary discoveries and contributions. Section VIII subsequently provides a delineation of prospective paths for future research. In addition, a comprehensive compilation of references is provided to support and validate the research.

II. Related Studies

The authors of [5] used feature extraction techniques such as zoning, chain code, and Fourier descriptors to recognize Assamese handwritten numerals. Table I presents a compilation of the relevant literature pertaining to ensemble methods. The extracted features could be used for word recognition. The study explores a deep learning-based approach for Assamese text recognition [6]. The results show that preprocessing can help a convolutional neural network (CNN) architecture recognize words accurately. Several studies have examined Assamese handwriting recognition issues. An ensemble system uses deep learning models like CNNs and LSTMs to improve recognition accuracy [7]. Ensemble techniques can improve Assamese handwriting recognition and social inclusion, according to this study. The paper proposes an adaptive approach for handwritten Assamese word recognition [8]. The horizontal, vertical, and gradient profiles of word images are used to extract features. The system uses a hybrid classifier that combines SVM and ANN benefits. This study contributes to inclusive word recognition by focusing on Assamese handwriting recognition challenges. This paper focuses on recognizing characters in offline Assamese handwriting [5]. The CNN architecture presented in this study is designed for Assamese character recognition. Word recognition systems require precise character identification, and this study improves Assamese word recognition. This study proposes a Hidden Markov Model (HMM) and Support Vector Machine (SVM) approach for Assamese online character recognition [9]. These models are also compared for efficacy. According to [10], an OCR system for handwritten Assamese characters uses Artificial Neural Networks (ANN) for character segmentation. Character segmentation is done using horizontal and vertical projection on handwritten text. In [11], researchers examine Academic literature proposes algorithms for online handwriting and machine-printed Assamese language text recognition. The Assamese language has more cursive writing than English and others. Feature selection (FS) extracts many features from simple to complex data.

Bagging and AdaBoost classifiers are popular ensemble learning methods for handwritten character recognition. Using different subsets of training data, many researchers have trained decision trees, support vector machines, and neural networks. The authors of [12] show that Bagging and AdaBoost Classifiers enhance devnagari document recognition accuracy. Researchers in [13] proved that statistical features can accurately classify handwritten Assamese language. Several studies have shown that deep learning techniques like CNNs, RNNs, and Deep Boltzmann Machines can effectively recognize handwritten text [14]. The above Assamese word recognition papers offer significant contributions and methods. They demonstrate the ongoing effort to improve social inclusivity by creating accurate and reliable recognition systems. Table IV presents a concise overview of the ensemble learning techniques employed in different scripts.

TABLE I. The Literature Concerning Ensemble Learning Methods Applied to Scripts

Year	Author	Technique	Dataset	Advantages
2007	Sarma[5]	Feature extraction (zoning, chain code, Fourier descriptors)	Handwritten Assamese numerals	Potential for word recognition
2007	Sarma (Character Recognition) [5]	CNN architecture for character recognition	Offline Assamese handwriting	Specific to characters
2013	Sarma[9]	HMM and SVM models	Assamese online characters	Utilizes multiple models
2015	Singh[8]	Hybrid classifier (SVMs, ANNs)	Handwritten Assamese words	Focuses on difficulties
2018	Jangid[14]	Deep learning techniques (CNNs, RNNs, etc.)	Handwritten text	Utilizes deep learning
2019	Alvear[6]	CNN architecture with preprocessing	Assamese text	Utilizes deep learning
2019	Narang[12]	Bagging and AdaBoost Classifiers	Devanagari documents	Ensemble learning
2019	Narang[13]	Statistical features	Handwritten Assamese language	Utilizes statistical features
2021	Choudhury [7]	Ensemble system (CNNs, LSTMs)	Handwritten Assamese text	Societal inclusion
2019	Chourasia [10]	ANN for character segmentation	Handwritten Assamese characters	Specific to character recognition
2022	Ghosh[11]	Various algorithms	Online handwriting and machine-printed Assamese text	Explores various algorithms

TABLE II. Summary of Instance Counts for Each Class Label

Sl. No.	Class Label	Instance Count
1	Burhi_aair_xadhu	2040
2	Bezbaruahr Rasanawali (Vol 2)	2065
3	Kirttana_and_Ghosha	2050
4	Gauburha	2062
5	Jilikoni	2057
	Total	10274

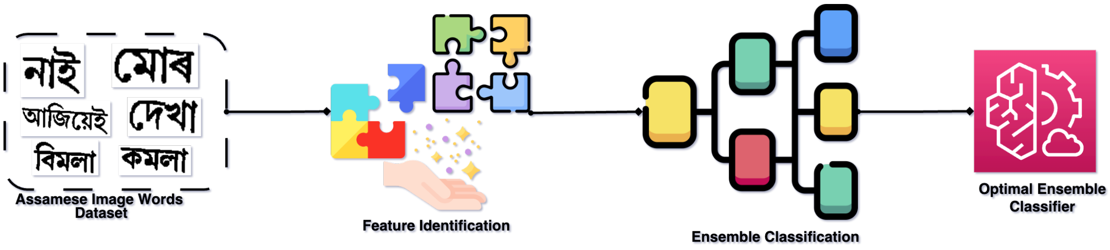


Fig. 1. The workflow of the methodology.

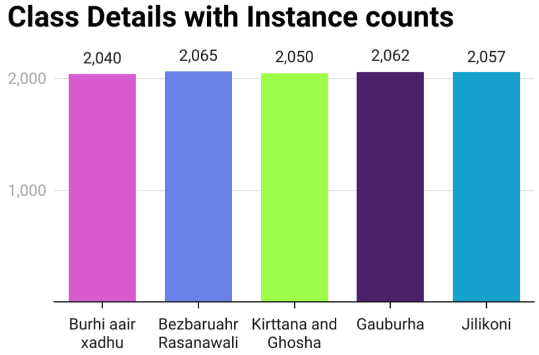


Fig. 2. Class details with instance counts.

III. Materials and Methods

A. Dataset Creation

The dataset curation approach for the experiments entailed a rigorous manual technique. The visual representations were created utilizing publicly accessible web resources related to the Jonaki and Shankari literary periods in Assamese literature. Access to the dataset can be provided upon a formal request, taking into account its potential significance.

The dataset used in this research was carefully chosen and organized. It consisted of photographs obtained from freely available online publications that were related to the Jonaki and Shankari periods in Assamese literature. The importance of studying these literary periods is in their contribution to the cultural heritage of Assam. The collection notably features images from five specific novels, namely, “Burhi Aair Xadhu”, “Bezbaruahr Rasanawali (Vol 2)”, “Kirttana and Ghosha”, “Gauburha”, and “Jilikoni”. The dataset, comprising images extracted from particular books, is presented in Table II, and Fig. 2 displays the specific information regarding the classes and the corresponding number of instances in our dataset.

Borah et al. offer a thorough and complete explanation of a detailed segmentation process [15]. The experimental phase utilized a dataset consisting of 10,274 photographs, which were classified into five unique categories, as detailed in Table II and Fig. 2.

B. Feature Engineering

Table III presents a comprehensive compilation of 1523 distinct characteristics used in the image dataset. These characteristics encompass a wide range of diverse attributes. Every attribute contributes to the creation of a comprehensive representation of handwritten word images. Significant metrics in this context encompass CHA (Convex Hull Area) and CHP (Convex Hull Perimeter), which provide valuable information about the geometric attributes of words. SPLBP (Spatial Layout Binary Pattern) provides valuable information regarding spatial layout attributes, including position, orientation, and scale. The analysis is improved by integrating features such as PHOG (Pyramid of Histograms of Oriented Gradients) and EHD (Edge Histogram Descriptors), which offer insights into texture, structure, and shape. The local shape properties are understood by considering additional characteristics such as kurtosis, rectangularity, volume, compactness, and HuMoments. The recognition system requires a wide variety of features in order to accurately distinguish and categorize handwritten Assamese words. This will improve the accuracy and robustness of ensemble classification methods. Additionally, [16] offers a comprehensive examination of various shape-based features utilized in CBIR.

TABLE III. Features Implemented on the Image Dataset

Feature Name	Count	Description
CHA	1	Area of smallest convex polygon containing the object.
CHP	1	Perimeter of smallest convex polygon containing the object.
Compactness	1	Measure of object’s packing density.
Contlength	1	Length of object contour or boundary.
EHD	80	Edge Histogram descriptors for shape analysis.
HuMoments	7	Moments computed from central moments for shape description.
Kurtosis	1	Measure of distribution’s “peakedness” or “flatness”.
MaAL	1	Length of object’s major axis.
MiAL	1	Length of object’s minor axis.
Num_corners	1	Number of corners or vertices in the object’s contour.
Num_holes	1	Number of holes or voids in the object.
Perimeter	1	Length of the object’s contour.
Rectangularity	1	Ratio of object area to minimum bounding rectangle area.
ShapeIndex	36	Scalar value characterizing local shape based on curvature.
Skewness	1	Measure of object’s distribution asymmetry.
Solidity	1	Ratio of object’s area to its convex hull area.
SPLBP	756	Features describing spatial layout: position, orientation, scale.
Volume	1	Volume or space occupied by the object.
PHOG	630	Pyramid of Histograms of Oriented Gradients
Total	1523	

C. Ensemble Classification

Ensemble classification is a commonly employed technique in the domain of machine learning that leverages the combined power of multiple models to enhance the precision of predictions [17]. In order to create an ensemble model that is more reliable and accurate, the proposed methodology integrates forecasts made by numerous base classifiers, also referred to as “weak learners”. Ensemble methods, including bagging, boosting, and stacking, are employed to introduce diversity among individual models and collectively mitigate their respective limitations. An illustration of a technique employed in the field of machine learning is bagging, wherein multiple instances of base models are generated on bootstrapped samples. This particular methodology successfully decreases variance and addresses the potential problem of overfitting. By giving weights to samples that were wrongly classified, the boosting algorithm makes it possible for the model's performance to be improved over and over again. The stacking technique entails the amalgamation of multiple models through the utilization of a meta-learner, with the aim of capitalizing on their respective strengths. Ensemble classification is extensively employed across diverse domains, including image recognition, natural language processing, and financial forecasting, with the primary aim of achieving enhanced predictive accuracy. The methodology employed in our study is depicted in Fig. 1, outlining the step-by-step process of improving Assamese word recognition.

1) Gradient Boosting (GB): The technique of Gradient Boosting is employed to optimize a loss function through the sequential addition of weak learners, typically in the form of decision trees [18]. The ultimate forecast is computed by aggregating the individual predictions of the learners using a weighted sum.

Mathematically, in each iteration, we update the model as follows:

$$F_t(x) = F_{t-1}(x) + \arg\min_h \left(\sum_{i=1}^{N} L(y_i, F_{t-1}(x_i) + h(x_i)) \right) \tag{1}$$

Definitions:

1. $F_t(x)$: The ensemble's prediction at iteration t. 2. $h(x)$: The weak learner's prediction. 3. $L(y_i, F(x_i))$: The loss function, typically squared error for regression or cross-entropy for classification.

2) CatBoost (CB): In order to enhance the training procedure, CatBoost employs the ordered boosting technique, which takes into account the ordering of categorical variables [19]. This methodology facilitates the preservation of the inherent hierarchy of categorical attributes, which can prove advantageous in a multitude of contexts including recommendation and ranking systems. By integrating the natural order of categorical variables, CatBoost has the capability to augment the model's predictive performance.

CatBoost, apart from employing ordered boosting, incorporates a statistical technique to mitigate the risk of overfitting in the context of categorical data. Overfitting occurs when a model learns an excessive amount of the training data, including any noise or random fluctuations, which can result in inadequate generalization to new data. CatBoost implements a regularization technique to mitigate the risk of overfitting, specifically in the context of categorical features.

From a mathematical standpoint, it can be observed that this approach optimizes the identical loss function as gradient boosting. However, it distinguishes itself by employing distinct methods to handle categorical features.

$$\text{CB} = \sum_{i=1}^{N} L(y_i, F(x_i)) + \sum_{j=1}^{J} \Omega(C_j) \tag{2}$$

Where: - $L(y_i, F(x_i))$ is the loss function that measures the difference between the predicted values and the true labels.
- $F(x_i)$ represents the model's prediction for the i-th data point.
- C_j represents categorical features and $\Omega(C_j)$ is a regularization term applied specifically to categorical features.
- J is the total number of categorical features.

3) Random Forest (RF): Random Forest is a machine learning algorithm that operates by aggregating multiple decision trees. As such, it does not possess a singular equation that fully encompasses its functionality. The RF algorithm creates a collection of decision trees and aggregates their predictions using either voting (for classification [20], [21]) or averaging (for regression [22]).

4) XGBoost: XGBoost effectively optimizes the given objective function in order to construct an ensemble of decision trees, rendering it a robust and efficient algorithm that is extensively employed in machine learning competitions.

$$\text{XGBoost} = \sum_{i=1}^{N} L(y_i, F(x_i)) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

Where:
-L(y_i, F(x_i)) & is the loss function.
-Omega(f_k) & is the regularization term for each tree.

5) Bagging: Bagging improves decision tree-based predictive models' accuracy and resilience. Original model is a “weak learner”, usually a decision tree. Bootstrap samples create multiple instances by training the base model on different training data subsets. This subset is created by random sampling with replacement. Academic literature calls these subsets “bootstrap samples”. Multiple base models trained on bootstrap samples are exposed to slightly different dataset variations. Predicting future events. Model predictions are aggregated. Regression aggregation averages predictions, while classification determines majority vote. Bagging reduces variance, overfitting, and generalization by using trained model heterogeneity. Thus, it enhances prediction.

The Bagging prediction can be represented as:

$$\text{Bagging Prediction} = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \tag{4}$$

Where: - Bagging Prediction is the final prediction made by Bagging. - N is the number of base models (often decision trees) created through bootstrapped samples. - $f_i(x)$ represents the prediction of the i-th base model on input data x.

6) AdaBoost: AdaBoost, also known as Adaptive Boosting, is a machine learning algorithm that employs the technique of ensemble learning to construct a robust model by aggregating multiple weak learners [23]. The algorithm employs an iterative process whereby distinct weights are assigned to each weak learner in accordance with their respective performance. The primary objective is to rectify the errors made by preceding models. The ultimate forecast is derived by calculating a weighted sum of the prognostications made by these inferior learners.

The AdaBoost algorithm can be mathematically summarized as follows:

1) **Initialization:** Start by initializing the sample weights w_i uniformly, where i ranges from 1 to the number of training samples.
2) **Iteration t:**
   a) Train a weak learner, denoted as $h_t(x)$, on the training data with the current sample weights w_i.
   b) Compute the weighted error ϵ_t of $h_t(x)$ on the training data:
   $$\epsilon_t = \sum_{i=1}^{N} w_i \cdot I(y_i \neq h_t(x_i)) \tag{5}$$
   where N is the number of training samples, y_i is the true label, x_i is the input data, and I is the indicator function.
   c) Compute the importance weight of $h_t(x)$:
   $$\alpha_t = \frac{1}{2} \cdot \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{6}$$
   d) Update the sample weights for the next iteration:
   $$w_{i,t+1} = w_i \cdot \exp\left(-\alpha_t \cdot y_i \cdot h_t(x_i)\right) \tag{7}$$
   Normalize the weights so that they sum up to 1:
   $$w_{i,t+1} = \frac{w_{i,t+1}}{\sum_{i=1}^{N} w_{i,t+1}} \tag{8}$$
3) Repeat the above steps for a predefined number of iterations or until a stopping criterion is met.
4) The final prediction $F(x)$ is obtained by combining the predictions of weak learners with their importance weights:
$$F(x) = \sum_{t=1}^{T} \alpha_t \cdot h_t(x) \tag{9}$$

7) Stacking: Stacking combines multiple base models by training a meta-model on their predictions. The meta-model learns to weigh the predictions of the base models optimally [25].

Mathematically, stacking can be represented as follows:

$$\hat{y} = g(f_1(x), f_2(x), \ldots, f_k(x)) \tag{10}$$

Where:

$\hat{y}$ is the final prediction.

$f_i(x)$ are the predictions of individual base models.

g is the meta-model, which can be a LR, DT etc.

8) Histogram-based Gradient Boosting: Histogram-Based Gradient Boosting uses histograms to speed up training [24]. It optimizes the same loss function as traditional gradient boosting but employs histogram-based techniques for better efficiency.

The mathematical details of HistGradientBoosting involve optimizing the loss function similar to gradient boosting but with histogram-specific optimizations.

The optimization objective of Histogram-Based Gradient Boosting (HistGradientBoosting) can be summarized as follows:

$$\min_{F(x)} \sum_{i=1}^{N} L(y_i, F(x_i)) + \sum_{j=1}^{J} \Omega(C_j) \tag{11}$$

Where: - $\min_{F(x)}$ denotes the minimization of the objective with respect to the ensemble model $F(x)$. - N is the number of training samples. - $L(y_i, F(x_i))$ is the loss function that measures the difference between the true labels y_i and the predictions $F(x_i)$. - J represents the categorical features, and $\Omega(C_j)$ is a regularization term specific to categorical features.

This equation provides a simplified representation of the optimization objective in HistGradientBoosting, highlighting the key components involved in gradient boosting with histogram-based techniques.

TABLE IV. SUMMARY OF ENSEMBLE MODELS ASSAMESE WORD RECOGNITION

Model	Description
Gradient Boosting (GB)	Gradient Boosting is an ensemble learning method that builds multiple decision trees sequentially. Each tree corrects the errors of the previous one. It's a powerful model for classification and regression tasks, known for its high predictive accuracy.
CatBoost (CB)	CatBoost is a gradient boosting algorithm that is particularly effective for categorical feature handling. It automatically handles categorical data, reducing the need for preprocessing, and can work well with both numerical and categorical features.
Random Forest (RF)	Random Forest is an ensemble of decision trees. It builds multiple trees and combines their predictions through voting (classification) or averaging (regression). It's robust, handles high-dimensional data well, and is less prone to overfitting.
XGBoost	XGBoost (Extreme Gradient Boosting) is an efficient gradient boosting algorithm known for its speed and performance. It's highly customizable and widely used in machine learning competitions due to its predictive power.
Bagging	Bagging stands for Bootstrap Aggregating. It's an ensemble method that builds multiple instances of a base model (usually decision trees) on bootstrapped samples of the data. It reduces variance and helps in avoiding overfitting.
AdaBoost	AdaBoost (Adaptive Boosting) is another boosting algorithm that focuses on the weaknesses of the base model. It assigns weights to misclassified samples and combines multiple weak learners to create a strong ensemble model.
Stacking	Stacking, or Stacked Generalization, combines multiple base models by training a meta-model on their predictions. It leverages the strengths of different models, potentially improving overall performance.
HistGradientBoosting	Histogram-Based Gradient Boosting is an efficient variant of gradient boosting that uses histograms to speed up training. It's particularly useful for large datasets and high-dimensional data.

IV. PERFORMANCE METRICS

In the domain of machine learning evaluation, a variety of fundamental performance metrics are frequently utilized to

TABLE V. PERFORMANCE METRICS FOR VARIOUS MACHINE LEARNING MODELS

Model	Accuracy	Precision	Recall	Kappa	F1-score	MCC	Build Time (s)	Run Time (s)	AUC
GB	0.9603	0.9603	0.9603	0.9504	0.9603	0.9504	142.4156	0.0255	0.9972
CB	0.9770	0.9771	0.9770	0.9713	0.9770	0.9713	27.2524	0.5145	0.9984
RF	0.9366	0.9371	0.9366	0.9207	0.9366	0.9208	2.2948	0.0403	0.9943
XGBoost	0.9751	0.9751	0.9751	0.9689	0.9751	0.9689	15.5116	0.0125	0.9989
Bagging	0.8828	0.8829	0.8828	0.8535	0.8827	0.8536	8.0071	0.0876	0.9786
AdaBoost	0.7777	0.7831	0.7777	0.7222	0.7785	0.7233	6.0609	0.1455	0.9271
Stacking	0.9786	0.9787	0.9786	0.9732	0.9786	0.9733	796.6909	4.9693	0.9988
HistGB	0.9743	0.9744	0.9743	0.9679	0.9743	0.9679	36.7397	0.0888	0.9990

TABLE VI. EXISTING VS PROPOSED FEATURE-BASED METHODS

Authors	Scripts	Word Count	Feature Set	Classifier	Accuracy (%)
Shaw and Parui [26]	Devanagari	13,000	Stroke based (Stage-1); Wavelet (Stage-2)	HMM (Stage-1);	91.25
Singh et al. [27]	Devanagari	28,500	Curvelet transform	SVM and KNN	85.6 (SVM); 93.21 (KNN)
Singh[28]	Devanagari	20,000	Combination of uniform zoning, diagonal and centroid features	Gradient boosted decision tree	94.33
Malakar et al. [29]	Hindi	4,620	Low-level features	MLP	96.82
Kaur and Kumar [30]	Gurumukhi	40,000	Zoning features	XGBoost	91.66
Ghosh et al. [31]	Bangla	7,500	Gradient features and modified SCF; MA-based wrapper filter selection approach	MLP	93
Malakar et al.[32], [33]	Bangla	12,000	Gradient-based and elliptical	MLP	95.3

Bhunia et al. [34]	Bangla, Devanagari, Gurumukhi	3,856;
		3,589;
		3,142
			PHOG feature	HMM (Middle-zone), SVM (Upper/Lower zone)	>60

Proposed method (Feature-Based Ensemble)	Assamese	10274	Combination of Multiple Low level - Shape, Region, Descriptor-Based Features,	CatBoost, XGBoost	97.7 (CB), 96.0 (XGBoost)

evaluate the efficacy of classification models. These metrics offer significant insights into the predictive capabilities and overall quality of a model. In this section, we present a concise summary of several key performance metrics.

A. Accuracy

Accuracy is a crucial metric that quantifies the ratio of accurately classified instances to the total number of instances in the dataset. The aforementioned statement offers a comprehensive evaluation of the accuracy of the model's predictions on a global scale.

B. Precision

Precision is a metric that quantifies the ratio of correctly predicted positive instances to the total number of positive predictions made by the model. The evaluation assesses the model's capacity to minimize the occurrence of false positive predictions.

C. Recall

The term "recall", which is also referred to as sensitivity or true positive rate, is a metric used to measure the proportion of true positive predictions in relation to all the actual positive instances. The evaluation measures the model's capacity to accurately identify and include all pertinent positive instances.

D. Kappa

The Cohen's Kappa statistic serves as a quantitative measure to assess the level of agreement between the predictions generated by a model and the observed outcomes. The method takes into consideration the potential occurrence of coincidental agreement and offers an indication of the model's efficacy beyond what would be expected by random chance.

E. F1-score

The F1-score can be defined as the mathematical average of precision and recall, specifically calculated using the harmonic mean. The provided analysis presents a comprehensive evaluation of the model's capacity to effectively balance precision and recall, making it particularly advantageous in scenarios involving imbalanced datasets.

F. Matthews Correlation Coefficient (MCC)

The Matthews Correlation Coefficient (MCC) is a metric used to evaluate the degree of correlation between the predictions made by a model and the actual labels assigned to the data. It takes into account all four categories: true positives, true negatives, false positives, and false negatives. The aforementioned metric offers a thorough evaluation of the performance of the model.

G. Build Time (s) and Run Time (s)

The term "Build Time" denotes the duration, measured in seconds, necessary for the training or construction of a machine learning model. The metric of Run Time quantifies the duration, expressed in seconds, required for generating predictions on novel and unobserved data. The metrics presented in this context serve as indicators of the computational efficiency of the model.

H. AUC Area Under the ROC Curve (AUC)

The quantification of a model's ability to differentiate between positive and negative classes is accomplished by the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). A greater Area Under the Curve (AUC) value signifies an enhanced capacity to differentiate between different classes.

Performance metrics play a crucial role in the assessment and comparison of the efficacy of machine learning models in classification tasks.

I. Training, Testing and Validation Details

The dataset was divided into training and testing sets using a 70-30 split, allowing for a comprehensive evaluation of ensemble classification techniques for identifying Assamese words. More precisely, 70% of the data was allocated for training and denoted by the variables X_{train} and y_{train}. The remaining 30% of the data, referred to as X_{test} and y_{test}, was used for testing. The division was performed using the `train_test_split` function, with the `test_size` parameter set to 0.3. In addition, an additional 20% of the training set was allocated for validation purposes. The validation subset played a vital role in optimizing and validating the models. It enabled the adjustment of parameters and the prevention of overfitting, ensuring the best possible performance of the model on new, unseen data. The results were greatly impacted by the implementation of the data segmentation strategy, which offered a thorough approach for training and evaluating the model. This, in turn, improved the reliability and accuracy of the findings.

J. Experimental Environment Details

The study was conducted on a computational system in the experimental environment, which had the following specifications. The system utilized Python version 3.8.17 and functioned on the Darwin operating system with Kernel Version 23.2.0. The system architecture was 64-bit, with a RAM capacity of 32.00 GB. The CPU was equipped with a configuration consisting of 10 cores and 10 threads. The hardware and software specifications enhance the transparency and reproducibility of the experimental setup, guaranteeing a strong basis for the study's results.

V. Results

The present study provides a comprehensive evaluation of various machine learning models within the framework of a classification task. Each model is assessed using a variety of performance metrics, including accuracy, precision, recall, Kappa, F1-score, MCC (Matthews Correlation Coefficient), build time, run time, and AUC (Area Under the ROC Curve). Testing how well a model works involves checking how well it correctly labels instances, how well it balances precision and recall, how well it matches real-world results, and how quickly it can do the calculations (see Table V). It is noteworthy to mention that specific models, such as CatBoost (CB) and XGBoost, demonstrate a significant degree of accuracy and AUC values, indicating their strong discriminatory abilities. In contrast, AdaBoost demonstrates reduced accuracy and Matthews Correlation Coefficient (MCC) values, suggesting the existence of potential areas for improvement as seen in Fig. 3 and 4a. The table serves as a valuable instrument for selecting the most suitable machine learning model based on specific classification requirements, considering both predictive accuracy and computational efficiency. Fig. 3 illustrates the performance of the classifiers on our dataset, specifically emphasizing accuracy and precision and Fig. 4a and 4b provides a performance analysis of the models, showcasing their ROC values and computational efficiency.

VI. Discussion

This study offers a comprehensive assessment of machine learning models in the context of a classification task. This analysis yields significant insights of note.

The CatBoost (**CB**) model demonstrates superior performance across various metrics, establishing its dominance in the field. The model demonstrates exceptional performance in terms of accuracy (97.7%), precision, and recall, highlighting its proficiency in making precise predictions and effectively capturing positive instances.

The competence of Gradient Boosting (**GB**) is notable, as it consistently demonstrates high accuracy (96.0%) and performs competitively across multiple metrics. It is a highly suitable option for precise classification tasks.

The achievement of high F1-scores, which effectively balance precision and recall, is notable in the performance of **CB** and **XGBoost**. This particular attribute holds significant value when addressing datasets that exhibit imbalances.

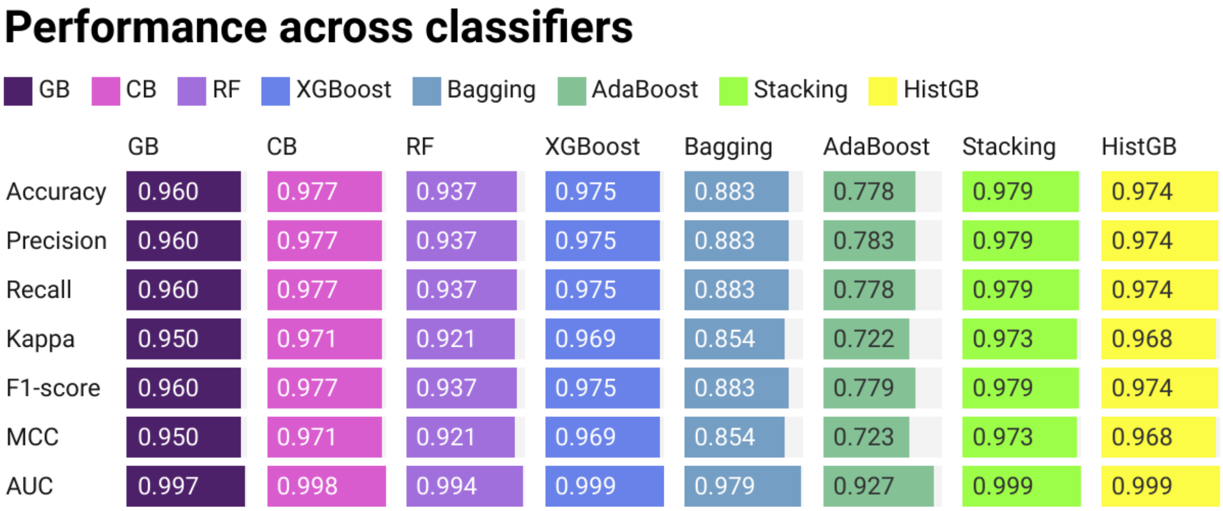
The observed Kappa values for **CB** and **Stacking** indicate a significant level of agreement that surpasses what would be expected by chance alone. This suggests their credibility in generating forecasts that surpass mere chance concurrence.

The discriminatory power of all models is exceptional, as evidenced by their AUC values that approach 1. This suggests their proficiency in effectively discerning between positive and negative classes.

Computational efficiency is a crucial aspect to consider when evaluating the performance of algorithms such as **CB** and **XGBoost**. Although these algorithms exhibit exceptional performance, it is important to note that they necessitate a greater allocation of computational resources. In contrast, **AdaBoost** and **Random Forest** algorithms provide expedited construction and execution durations, rendering them appropriate for situations where computational efficiency is of paramount importance.

A. Perfomance with Reference to other Script based Works

The data as seen in Table VI is a comparative analysis of the limited number of techniques utilized in the field of document recognition and analysis across a variety of scripts. The methodologies being evaluated are attributed to distinguished researchers who have implemented unique approaches for the extraction and categorization of features. The comparative analysis in question examines a wide range of scripts, which comprise Assamese, Devanagari, Hindi, Gurumukhi, and Bangla.

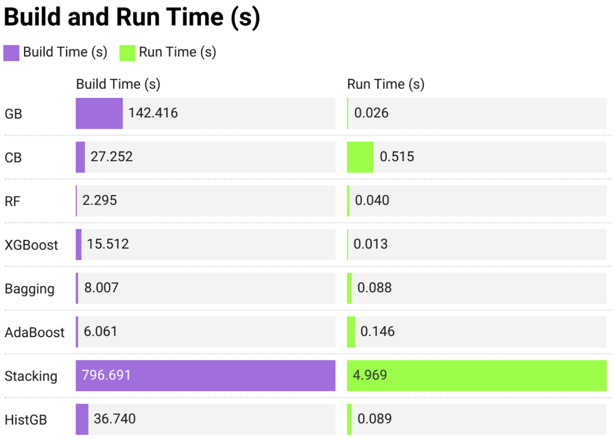


Performance across classifiers

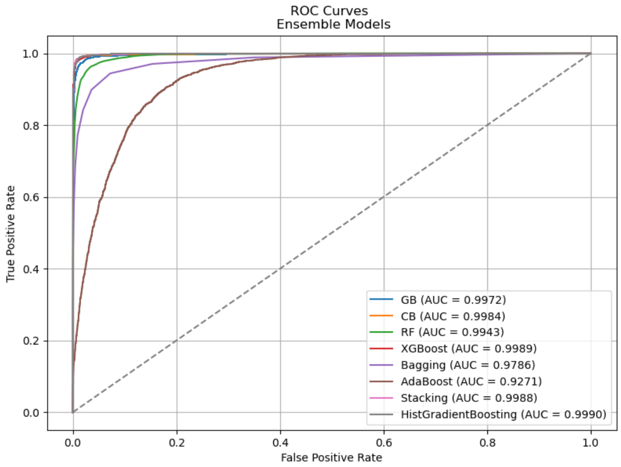
🟪 GB 🟪 CB 🟪 RF 🟦 XGBoost 🟦 Bagging 🟩 AdaBoost 🟩 Stacking 🟨 HistGB

	GB	CB	RF	XGBoost	Bagging	AdaBoost	Stacking	HistGB
Accuracy	0.960	0.977	0.937	0.975	0.883	0.778	0.979	0.974
Precision	0.960	0.977	0.937	0.975	0.883	0.783	0.979	0.974
Recall	0.960	0.977	0.937	0.975	0.883	0.778	0.979	0.974
Kappa	0.950	0.971	0.921	0.969	0.854	0.722	0.973	0.968
F1-score	0.960	0.977	0.937	0.975	0.883	0.779	0.979	0.974
MCC	0.950	0.971	0.921	0.969	0.854	0.723	0.973	0.968
AUC	0.997	0.998	0.994	0.999	0.979	0.927	0.999	0.999

Fig. 3. The performance of several classifiers on the dataset.



(a) Build and run times of several classifiers.



(b) ROC values for classifiers.

Fig. 4. Performance analysis of models.

The methods described herein employ datasets with substantial variation in word count, which spans from 3,142 to 40,000. This discrepancy is indicative of the vast and varied scale of the document corpora under investigation. Prominent feature extraction methodologies include modified spatial co-occurrence functions (SCF), stroke-based techniques, curvelet transforms, combinations of zoning and centroid features, low-level features, zoning features, and gradient features. Moreover, the integration of ensemble techniques, exemplified by the proposed method, encompasses a multitude of low-level characteristics, including those based on shape, region, and descriptors.

Hidden Markov Models (HMM), Support Vector Machines (SVM), k-Nearest Neighbors (KNN), gradient-boosted decision trees, Multi-Layer Perceptrons (MLP), XGBoost, and CatBoost are among the classifiers utilized in the assessed methodologies. The aforementioned classifiers demonstrate an extensive array of algorithmic methodologies, which accurately represents the intricate demands of document analysis across various scripts.

One crucial metric for evaluating the effectiveness of the suggested methodologies is accuracy. The range of obtained accuracy values is between 85.6% and 97.7%. The performance of the proposed method, which employs a feature-based ensemble approach for Assamese script recognition, is noteworthy. By attaining an accuracy of 97.7% with CatBoost and 96.0% with XGBoost, this ensemble method solidifies its position as a formidable competitor within the realm of script recognition methodologies.

VII. Conclusion

The investigation of Assamese word recognition demonstrated superior performance in terms of F1-scores, accuracy, precision, and recall. This was achieved by utilizing ensemble methods and feature extraction techniques, with a specific focus on the effectiveness of CatBoost and XGBoost.

The research inquiry has emphasized the findings regarding the efficacy of different ensemble methods, particularly in the identification of Assamese words. Furthermore, it is emphasized that although AdaBoost and Random Forest can be effective alternatives, especially in scenarios with limited computational resources, they demonstrate slightly lower performance metrics compared to CatBoost and XGBoost.

The methodology's resilience and practicality are showcased by employing a comprehensive dataset comprising 10,000 words and diverse feature extraction methodologies.

A significant advancement in the field of computational linguistics has been achieved by successfully developing a method for recognizing Assamese words. This has resulted in the promotion of technological diversity across different languages.

VIII. Future Work

This study presents various opportunities for future research. An important focus is the incorporation of deep learning methods to improve the process of extracting distinctive characteristics and accurately categorizing Assamese words. Investigating recurrent neural networks and convolutional neural networks has the potential to yield substantial enhancements. Moreover, augmenting the dataset to encompass a wider range of handwriting styles and integrating multi-script recognition systems would significantly bolster the model's resilience. Finally, exploring the practical and influential application of these models in real-time on mobile devices and web applications would be a worthwhile direction.

IX. Conflict of Interest

The authors declare that there is no conflict of interest.

References

[1] S. Mahanta, *Assamese*, *Journal of the International Phonetic Association*, vol. 42, no. 2, pp. 217–224, 2012, doi: 10.1017/s0025100312000096.

[2] G. Upadhye, U. Kulkarni, and D. Mane, *Improved model configuration strategies for Kannada handwritten numeral recognition*, *Image Analysis & Stereology*, vol. 40, no. 3, pp. 181-191, 2021, doi: 10.5566/ias.2586.

[3] Y. Wen and R. Filik, *Electrophysiological dynamics of Chinese phonology during visual word recognition in Chinese-English bilinguals*, *Scientific Reports*, vol. 8, no. 1, 2018, doi: 10.1038/s41598-018-25072-w.

[4] S. Yuan, Y. Wei, and D. Zhao, *Computer-aided lung nodule recognition by SVM classifier based on combination of random undersampling and SMOTE*, *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 1-13, 2015, doi: 10.1155/2015/368674.

[5] K. K. Sarma, *MLP-based Assamese Character and Numeral Recognition using an Innovative Hybrid Feature Set*, in *IICAI*, December 2007, pp. 585-600, doi: 10.1109/IICAI.2007.357.

[6] R. F. Alvear-Sandoval, J. L. Sancho-Gómez, and A. R. Figueiras-Vidal, *On improving CNNs performance: The case of MNIST*, *Information Fusion*, vol. 52, pp. 106-109, 2019, doi: 10.1016/j.inffus.2019.03.016.

[7] A. Choudhury and K. K. Sarma, *A CNN-LSTM based ensemble framework for in-air handwritten Assamese character recognition*, *Multimedia Tools and Applications*, pp. 1-36, 2021, doi: 10.1007/s11042-021-11572-4.

[8] P. K. Singh, R. Sarkar, and M. Nasipuri, *Word-Level Script Identification Using Texture Based Features*, *International Journal of System Dynamics Applications (IJSDA)*, vol. 4, no. 2, pp. 74-94, 2015, doi: 10.4018/ijsda.2015040105.

[9] B. Sarma, K. Mehrotra, R. K. Naik, S. Raj Prasanna, S. Belhe, and C. Mahanta, *Handwritten Assamese Numeral Recognizer using HMM & SVM Classifiers*, in *2013 National Conference on Communications (NCC)*, February 2013, pp. 1-5, doi: 10.1109/NCC.2013.6487976.

[10] C. K. Chourasia and M. Barman, *Handwritten Assamese Character Recognition*, in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, March 2019, pp. 1-6, doi: 10.1109/I2CT45656.2019.9033947.

[11] T. Ghosh, S. Sen, S. M. Obaidullah, K. C. Santosh, K. Roy, and U. Pal, *Advances in Online Handwritten Recognition in the Last Decades*, *Computer Science Review*, vol. 46, 2022, pp. 100515, doi: 10.1016/j.cosrev.2022.100515.

[12] S. R. Narang, M. K. Jindal, and M. Kumar, *Devanagari Ancient Character Recognition Using DCT Features with Adaptive Boosting and Bootstrap Aggregating*, *Soft Computing*, vol. 23, 2019, pp. 13603-13614, doi: 10.1007/s00500-019-03973-2.

[13] S. Narang, M. K. Jindal, and M. Kumar, *Devanagari Ancient Documents Recognition Using Statistical Feature Extraction Techniques*, *Sādhanā*, vol. 44, 2019, pp. 1-8, doi: 10.1007/s12046-018-0997-7.

[14] M. Jangid and S. Srivastava, *Handwritten Devanagari Character Recognition Using Layer-Wise Training of Deep Convolutional Neural Networks and Adaptive Gradient Methods*, *Journal of Imaging*, vol. 4, no. 2, 2018, pp. 41, doi: 10.3390/jimaging4020041.

[15] N. Borah, U. Baruah, T. R. Mahesh, V. V. Kumar, D. R. Dorai, and J. Rajkumar Annad, *Efficient Assamese Word Recognition for Societal Empowerment: A Comparative Feature-Based Analysis*, *IEEE Access*, vol. 11, 2023, pp. 82302-82326, doi: 10.1109/ACCESS.2023.3301564.

[16] N. Borah and U. Baruah, *Feature Extraction Techniques for Shape-Based CBIR—A Survey*, in *Contemporary Issues in Communication, Cloud and Big Data Analytics*, H. K. D. Sarma, V. E. Balas, B. Bhuyan, and N. Dutta (Eds.), Lecture Notes in Networks and Systems, vol. 281, Springer, 2022, pp. 205-214, doi: 10.1007/978-981-16-4244-9_16.

[17] S. Priya, A. Agarwal, C. Ward, T. Locke, V. Monga, and G. Bathla, "Survival Prediction in Glioblastoma on Post-Contrast Magnetic Resonance Imaging Using Filtration-Based First-Order Texture Analysis: Comparison of Multiple Machine Learning Models," *The Neuroradiology Journal*, vol. 34, no. 4, pp. 355-362, 2021. doi: 10.1177/1971400921990766

[18] B. Fernandes, A. González-Briones, P. Novais, M. Calafate, C. Analide, and J. Neves, "An Adjective Selection Personality Assessment Method Using Gradient Boosting Machine Learning," *Processes*, vol. 8, no. 5, p. 618, 2020. doi: 10.3390/pr8050618

[19] L. Prokhorenkova, G. Gusev, A. Vorobev, A. Dorogush, and A. Gulin, "CatBoost: Unbiased Boosting with Categorical Features," arXiv preprint arXiv:1706.09516, 2017. doi: 10.48550/arxiv.1706.09516

[20] B. Balnarsaiah, T. Prasad, and P. Laxminarayana, "Pixel-Based SAR Image Classification Using Random Forest Algorithm," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 10, pp. 4351-4356, 2019. doi: 10.35940/ijitee.j9873.0881019

[21] D. Chutia, N. Borah, D. Baruah, et al., *An Effective Approach for Improving the Accuracy of a Random Forest Classifier in the Classification of Hyperion Data*, *Applied Geomat*, vol. 12, pp. 95–105, 2020, doi: 10.1007/s12518-019-00281-8.

[22] A. Jog, A. Carass, S. Roy, D. Pham, and J. Prince, "Random Forest Regression for Magnetic Resonance Image Synthesis," *Medical Image Analysis*, vol. 35, pp. 475-488, 2017. doi: 10.1016/j.media.2016.08.009

[23] A. Lykov, S. Muzychka, and K. Vaninsky, "The AdaBoost Flow," *Communications on Pure and Applied Mathematics*, vol. 68, no. 5, pp. 865-886, 2014. doi: 10.1002/cpa.21555

[24] M. Kashifi and I. Ahmad, "Efficient Histogram-Based Gradient Boosting Approach for Accident Severity Prediction with Multisource Data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2676, no. 6, pp. 236-258, 2022. doi: 10.1177/03611981221074370

[25] O. Petinrin and F. Saeed, "Stacked Ensemble for Bioactive Molecule Prediction," *IEEE Access*, vol. 7, pp. 153952-153957, 2019. doi: 10.1109/access.2019.2945422

[26] B. Shaw and S. K. Parui, "A two-stage recognition scheme for offline handwritten Devanagari words," In: Machine Interpretation of Patterns: Image Analysis and Data Mining, pp. 145-165, World Scientific, 2010.

[27] B. Singh, A. Mittal, M. Ansari, and D. Ghosh, ”Handwritten Devanagari word recognition: a curvelet transform based approach,” *International Journal of Computer Science and Engineering*, vol. 3, no. 4, pp. 1658-1665, 2011.

[28] S. Singh, N. K. Garg, and M. Kumar, ”On the performance analysis of various features and classifiers for handwritten Devanagari word recognition,” *Neural Computing and Applications*, vol. 35, no. 10, pp. 7509-7527, 2023.

[29] S. Malakar, P. Sharma, P.K. Singh, M. Das, R. Sarkar, and M. Nasipuri, ”A holistic approach for handwritten Hindi word recognition,” *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol. 7, no. 1, pp. 59-78, 2017.

[30] H. Kaur and M. Kumar, ”Offline handwritten Gurumukhi word recognition using eXtreme gradient boosting methodology,” *Soft Computing*, vol. 25, no. 6, pp. 4451-4464, 2021.

[31] M. Ghosh, S. Malakar, S. Bhowmik, R. Sarkar, and M. Nasipuri, ”Feature selection for handwritten word recognition using memetic algorithm,” In: J. Mandal, P. Dutta, and S. Mukhopadhyay (eds), *Advances in Intelligent Computing*, Studies in Computational Intelligence, vol. 687, pp. 103-124, 2019.

[32] S. Malakar, M. Ghosh, S. Bhowmik, R. Sarkar, and M. Nasipuri, ”A GA based hierarchical feature selection approach for handwritten word recognition,” *Neural Computing and Applications*, vol. 32, no. 7, pp. 2533-2552, 2020.

[33] S. Malakar, S. Paul, S. Kundu, S. Bhowmik, R. Sarkar, and M. Nasipuri, ”Handwritten word recognition using lottery ticket hypothesis based pruned CNN model: a new benchmark on CMATERdb212,” *Neural Computing and Applications*, vol. 32, no. 18, pp. 15209-15220, 2020.

[34] A. K. Bhunia, P. P. Roy, A. Mohta, and U. Pal, ”Cross-language framework for word recognition and spotting of Indic scripts,” *Pattern Recognition*, vol. 79, pp. 12-31, 2018

A Hybrid Deep Learning Framework for Efficient Sentiment Analysis

Asish Karthikeya Gogineni, S Kiran Sai Reddy, Harika Kakarala, Yaswanth Chowdary Gavini,
M Pavana Venkat, Koduru Hajarathaiah, Murali Krishna Enduri
Algorithms and Complexity Theory Lab-Department of Computer Science and Engineering,
SRM University-AP, Amaravati, India

***Abstract*—In the era of Microblogging and the rapid growth of online platforms, an exponential rise is shown in the volume of data generated by internet users across various domains. Additionally, the creation of digital or textual data is expanding significantly. This is because consumers respond to comments made on social media platforms regarding events or products based on their personal experiences. Sentiment analysis is usually used to accomplish this kind of classification on a large scale. It is described as the process of going through all user reviews and comments that are discovered in product reviews, events, or similar sources in order to look for unstructured text comments. Our study examines how deep learning models like LSTM, GRU, CNN, and hybrid models (LSTM+CNN, LSTM+GRU, GRU+CNN) capture complex sentiment patterns in text data. Additionally, we study integrating BOW and TF-IDF as complementing features to improve model predictive power. CNN with RNNs consistently improves outcomes, demonstrating the synergy between convolutional and recurrent neural network architectures in recognizing nuanced emotion subtleties.In addition, TF-IDF typically outperforms BOW in enhancing deep learning model sentiment analysis accuracy.**

Keywords*—*Sentiment analysis; LSTM; GRU; Convolutional Neural Networks (CNNs); BOW; TF-IDF

I. Introduction

Using accurate and reliable methods, sentiment analysis aims to automatically determine a text's sentiment and extract meaningful information. Various methodologies, tactics, algorithms, and approaches are investigated in sentiment analysis to understand textual emotions [1]. Its main purpose is to automatically classify text into neutral, joyful, furious, sad, and other emotions in addition to positive and negative ones. Sentiment analysis is essential for evaluating consumer evaluations of products and services since it reveals customer sentiment. The system uses advanced natural language processing algorithms to categorize customer evaluations, comments, and ratings as favorable, negative, or neutral (see Fig. 1). This enables organizations swiftly assess client opinion of their products.

Sentiment analysis also examines feedback to find traits that elicit positive or negative responses. Text-based movie reviews may be automatically analyzed and classified using sophisticated computational approaches for sentiment analysis [2]. Filmmakers and consumers may better communicate and enjoy film via movie reviews. Sentiment analysis algorithms evaluate reviews' language, tone, and context to determine positive, negative, or neutral sentiment [3]. This strategy helps filmmakers and spectators understand crowd responses. Sentiment analysis gives filmmakers vital audience feedback. Positive comments help filmmakers identify their talents by highlighting audience preferences. Conversely, negative feelings indicate places for progress. This helps improve storyline, character development, and cinematography in future projects. Sentiment analysis helps viewers choose movies [4]. Potential viewers may determine whether a film suits them by reading reviews.

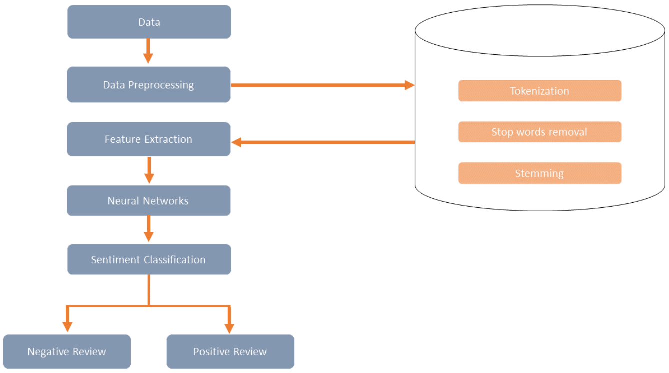


Fig. 1. Processing of DL algorithms.

An element of sentiment analysis employed in Coursera evaluations is the computerized evaluation of the emotion represented by students in their course remarks. Using ML or DL models trained on the labeled dataset, the sentiment analysis system learns to discover patterns that connect particular linguistic cues to particular sentiments [5]. The main motivation of this work is to analyze people's opinions by collecting reviews in the form of text and predicting the outcome, i.e., positive or negative reviews. Additionally, the goal is to assess the efficiency of deep learning models compared to machine learning techniques. When fresh Coursera reviews are fed into the algorithm, it forecasts the mood of each review, giving information about how students feel about the platform, instructors, course material, and more. This information can guide course improvements, influence instructional strategies, and offer valuable feedback to instructors, helping them refine their teaching methods based on learner sentiments.

The following sections offer a concise overview of pertinent details. In Section II, we examine prior research related to sentiment analysis, seeking to establish connections between existing knowledge and recent discoveries. Section III delves into the methodologies employed for this analysis, while Section IV is dedicated to detailing the dataset utilized in this paper. The conclusion of our work, along with the presentation of results, will bring this paper to a close.

A. Applications

1) Social media monitoring and brand management: Businesses use sentiment analysis to track mentions of their goods, providers of services, and brands on social media. By analyzing the sentiment of these mentions, you can gauge public perceptions, identify potential problems, and engage with your customers more effectively [6], [7].
2) Customer Opinion or Reviews: E-commerce platforms, restaurants, hotels, and other businesses use sentiment analysis to automatically process customer reviews. This helps us understand customer satisfaction, identify opportunities for improvement, and respond quickly to negative feedback [8], [9].
3) Movie and TV show review : Sentiment analysis is used to measure audience reactions to movies and TV shows, which aids in marketing and content creation [10], [11].
4) Product Development: Companies use sentiment analysis to analyze feedback and reviews of existing products, guiding improvements, and informing the development of new products that align with customer preferences [12], [13].

II. Related and Recent Work

In the recent past, the field of Machine Learning, Deep Learning, and Natural language processing (NLP) has witnessed a notable rise in sentiment analysis. With the huge expansion of textual information on the web, there is a growing need for automated tools that can accurately identify sentiment and emotional nuances within text. This increased focus on sentiment analysis has spurred the development of various novel methods and techniques that make use of deep learning models' capabilities to extract intricate sentiment insights from text data. These advances allow many applications, from social media sentiment analysis to corporate customer sentiment analysis. Here, we shall explain the recent contributions that have had a major influence on this subject and their relevance.

The work by Vateekul [14] used LSTM and DCNN deep learning algorithms on a Thai Twitter dataset. Deep learning outperformed several classic machine learning methods, according to their results. A reduced feature set and dimensionality reduction approaches were used in Akmal et al.'s clustering algorithm for emotional analysis [15]. The authors used deep learning to analyze sentiment in COVID-19 reviews [16], [17]. Their method uses an LSTM-RNN network and attention layers to improve features.For aspect extraction and emotion classification. A multitask learning technique was proposed by Akhtar et al. [18]. The extraction and sentiment categorization of aspects are seen to be separate activities. CNNs and BI-LSTMs execute the duties. The BI-LSTM layer infers Beginning and interior tags for each word using word embeddings to identify aspect keywords in a review sentence. Use a masking layer to remove non-aspect words from a text.

To extract sentiments from consumer evaluations about various characteristics of items, Pham and Le [19] used a multilayer architecture based on recurrent neural networks (RNNs). Their approach was used to examine 174,615 reviews relating to 1,768 hotels on tripadvisor.com. The results show that their technique has a lot of potential for sentiment analysis and hotel rating prediction. In this domain, Hassan and Mahmood used a hybrid technique that blends CNN and RNN architectures. They begin by training word embeddings with an unsupervised NLP model that has previously been fine-tuned on a large dataset. Following that, they use the capabilities of CNN for feature extraction and RNN for capturing interdependencies to detect sentiments in a variety of datasets [20].

We previously covered sentiment analysis using machine learning techniques [21] on similar datasets, as well as data cleaning and feature extraction, and this project serves as an expansion, digging into sentiment analysis using deep learning. Our findings reveal that machine learning models like Random Forest models demonstrate competitive performance, especially in scenarios with limited data. However, deep learning models, outperform Random Forest as the dataset size increases. We discuss the trade-offs between interpretability and performance, as Random Forest provides explicit feature importance, while deep learning models offer superior representation learning capabilities.

III. Methodology

This section offers a detailed overview of the approach employed for Sentimental Analysis using the neural networks.

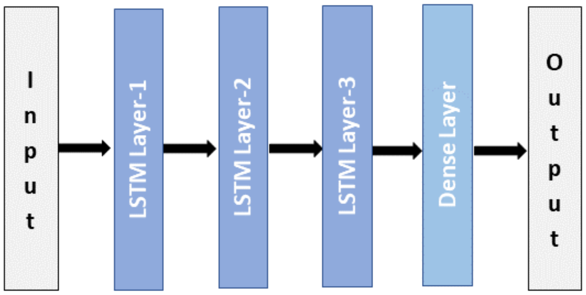


Fig. 2. LSTM Architecture.

A. LSTM

Long Short-Term Memory (LSTM) is a type of sophisticated recurrent neural network (RNN) crafted for proficient sequence modeling and prediction, mitigating the challenges of vanishing gradients in extended sequences [22]. In our model, first, we take input and pass it to LSTM Layers. Fig. 2 illustrates the three LSTM layers in our model. The first layer of the LSTM, Layer 1, has 32 units and returns sequences, passing on its output sequences to the subsequent layer. Layer 2, on the other hand, has 64 units and likewise returns sequences. Layer 3, on the other hand, has 128 units and does not return sequences, resulting in a fixed-size output. Data is eventually passed to a dense layer with 64 units and the ReLU activation function since it was an output layer for binary classification [23]. The output layer additionally has one neuron with a sigmoid activation function.

B. GRU

One type of specialised recurrent neural network (RNN) is the gated recurrent unit (GRU). The unique way that a GRU manages the temporal flow of information sets it apart from other RNN designs, such the Long Short-Term Memory (LSTM) network. The two gates that GRU uses to regulate information flow within the network are an update gate and a reset gate. This architectural simplicity of the GRU contributes to its ease of training and reduced susceptibility to overfitting, particularly in scenarios characterized by limited data [24]. The model that we used, contains three GRU layers as illustrated in Fig. 3. The first and second layer contains 32 and 64 units as well as return sequences, whereas the third layer contains 128 units with no return sequences. Subsequently, it is forwarded to a dense layer comprising 64 units, employing the ReLU activation function. The final layer is made up of a single neuron that has sigmoid activity, serving as the output layer for binary classification.

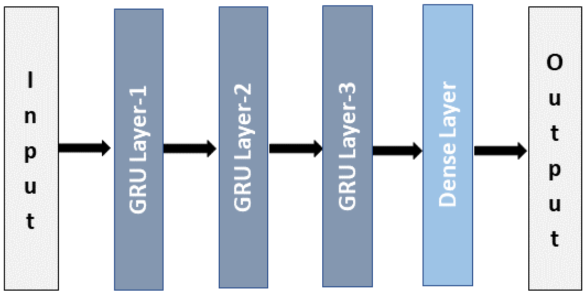


Fig. 3. GRU Architecture.

C. CNN

Convolutional neural networks, or CNNs for short, are a kind of Deep Learning neural network design that is frequently used in sentiment analysis, picture categorization, and other related applications. Convolutional neural networks, or CNNs, are known to be composed of several layers, including the input, convolutional, pooling, and dense layers. While the Pooling layer reduces the dimensions to lessen computational overhead, the Convolutional layer uses filters to extract features from the input. Ultimately, the fully connected layer is responsible for generating the final prediction [25]. In our model, the first layer is input layer which processes data with given shape and serves as entry point for review. Convolutional Layer 1, the next layer, has 32 filters, a 3×3 kernel, and ReLU activation. Using this, we then employed pooling. This pooling helps to reduce the size. Average and max poolings are the two main categories of pooling. Since max pooling performs better than average pooling and aids in the extraction of more significant features, we employed it with the default size. This process is repeated with Convolutional Layer 2, which has 64 filters, and then with an additional Convolutional Layer 3 using 128 filters. Convolutional layers extract various layers of information from the input. The main reason of using activation function as ReLU over other activation functions, i.e., Sigmoid, tanH and Softmax is that it does not activate all neurons at the same time. The output will be sent to the dense layer once the text has been processed via all of the pooling and convolutional layers. Since the dense layer requires input in the form of a 1-D array, we are unable to pass the convolutional layer's multi-dimensional output straight to it. To overcome this, we used the Flatten method between the dense layer and the convolutional layer. Finally, we used a dense layer to classify sentiment from convolutional layers as shown in Fig. 4. In our model, Since the output layer was intended for binary classification, it has one neuron with sigmoid activation function.

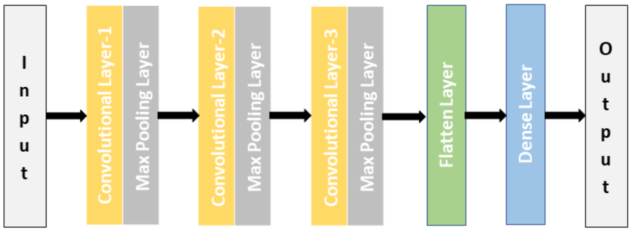


Fig. 4. CNN Architecture.

D. GRU+LSTM

As widely acknowledged, the Gated Recurrent Unit (GRU) is considered a highly promising algorithm within the realm of Recurrent Neural Networks (RNNs). In terms of functionality, both GRU and LSTM share similarities, However, GRU unifies the forget gate and input gate into a single update gate, using a single hidden state. Moreover, GRU creates a single state by merging the concealed state with the cell state. This efficient approach has led to GRU being recognized as a simplified variant of LSTM [26]. The model we have employed incorporates an initial layer with a single GRU layer containing 32 units and a return sequence. The subsequent layer features an LSTM layer with 64 units, also with a return sequence. We have added layer with a GRU and 128 filters the result then uses the ReLU activation algorithm to advance to a dense layer with 64 units. The final layer remains consistent with other hybrid models as depicted in Fig. 5.

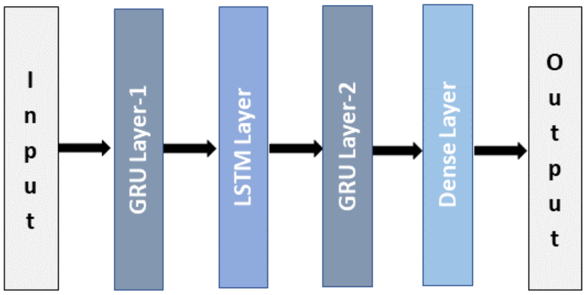


Fig. 5. GRU+LSTM Architecture.

E. CNN+LSTM

A Convolutional Neural Network (CNN) operates on sequential data by employing sliding convolutional filters over the input, enabling it to capture features from both spatial and temporal dimensions. In contrast, an LSTM network processes sequential data by iterative processing time steps, capturing long-term dependencies between them. A CNN-LSTM network combines convolutional and LSTM layers to effectively extract insights from training data [27]. In our model, The initial layer serves as the input layer, processing data with a specified shape. Following this, The first Convolutional Layer is then presented; it has 32 filters and uses a 3×3 kernel with activation as ReLU. After this operation, we apply pooling. This sequence is then replicated with Convolutional Layer 2, which incorporates 64 filters, and subsequently with an additional Convolutional Layer 3, utilizing 128 filters. Following the passage through the pooling and convolutional layers, the data is directed to the LSTM Layer, housing 64 units with a specific activation function. Subsequently, the output moves on to a 64-unit dense layer, employing the ReLU activation function. The final layer is composed of a 1 neuron utilizing a sigmoid activation function as illustrated in Fig. 6, which is ideal for binary classification, serving as the output layer.

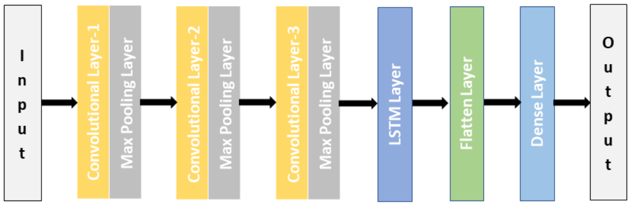


Fig. 6. CNN+LSTM Architecture.

F. GRU+CNN

The combination of CNN and Gated Recurrent Unit (GRU) components is our next hybrid model. The predictive performance of this dual-architecture method is enhanced. It also lessens the chance of overfitting and it is a novel work for the four different datasets in the context of sentimental analysis. The model we utilized, which is seen in Fig. 7, starts with a GRU (Gated Recurrent Unit) layer with 32 units and return sequences. Next, we add convolutional operations to the network by incorporating a Conv1D layer, which has 64 filters and a kernel size of 3. We next add another 128-unit GRU layer to the sequence, which is followed by a 64-unit dense layer with a ReLU activation function. Last but not least, we include a single-unit output layer and use a sigmoid activation function for binary classification.

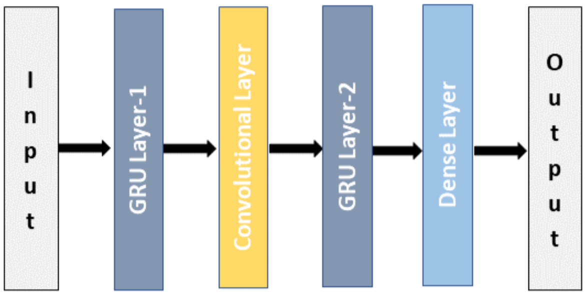


Fig. 7. GRU+CNN Architecture.

IV. Dataset Statistics

In our study, we harnessed a diverse array of four datasets to embark on comprehensive sentiment analysis. These datasets were meticulously selected for their relevance and diversity, thereby encompassing a broad spectrum of reviews from various domains. To elucidate further, the initial dataset centered around Movie Reviews and contained $34,000$ entries. Our second dataset, Coursera Reviews, proved to be expansive, comprising a total of $107,018$ reviews. Additionally, the Google Play Store Reviews dataset, our third one, provided valuable insights into the realm of mobile applications with its $12,485$ reviews. Lastly, the fourth dataset, Flipkart Reviews, encompassed a total of $9,976$ reviews, revealing a distinct polarity in emotional sentiment. This carefully curated assortment of datasets serves to ensure that our sentiment analysis model receives training from a diverse collection of linguistic expressions and contextual subtleties. There are $16,993$ instances of 0s and $17,007$ instances of 1s in the Movie Reviews data collection. There are $102,298$ occurrences of 0s and $4,720$ occurrences of 1s in the Coursera Reviews data set. There are $7,635$ instances of 0s and $4,850$ instances of 1s in Google Play Store Reviews. Finally, there are $8,975$ instances of 0s and $1,001$ instances of 1s in the Flipkart Reviews data set. Within their individual data sets, these binary values most likely indicate specific features or emotion labels connected with each review.

Formally, within a provided training dataset of reviews and their associated sentiment labels, a sentiment score of $'1'$ signifies a negative review, while a score of $'0'$ designates a positive one. We aim to characterize the feelings included within the supplied text collection appropriately.

A. Steps to Perform

In our research, we followed a structured process to prepare and preprocess our dataset [28].

- Data Cleaning: Special Character Removal: In this step, you remove special characters such as symbols, emojis, or any characters that don't contribute significantly to the text's meaning. This can be done using libraries like NLTK (Natural Language Toolkit) or regular expressions. Punctuation Removal: Similar to special characters, punctuation marks like periods, commas, and exclamation marks are often removed as they can add noise to the text.

- Tokenization: Dividing each sentence into individual words or tokens. For the majority of natural language processing (NLP) activities, this is a fundamental step.

- Stemming: The process of reducing words to their root or basic form is known as stemming. For instance, the terms “running,” “ran,” and “runs” may all have the same root, “run”. This can increase processing efficiency and aid in lowering the dimensionality of the text data.

- Feature Extraction: After that, we used Bag of Words and TF-idf (Term Frequency-Inverse Document Frequency) feature extraction methods to quantitatively represent text data for deep learning or machine learning models. TF-idf weighted words by importance in documents, while Bag of Words employed word counts to vectorize the text. Our data was better after these pretreatment stages and suitable for study analysis and modeling.

Bag-of-Words (BoW) characteristics were an important text analysis tool. BoW includes accumulating words to extract traits from text materials. For model training, a thorough vocabulary of different concepts from all training dataset records is needed. To facilitate text mining and information extraction, we used TF-IDF as a text data weight analysis approach. The significance of each word in a document is measured by TF-IDF. It normalizes word frequency according to text length, making it suitable for diverse document sizes. Additionally, TF-IDF rates each word's importance in a publication. Since it examines term frequency and normalizes it by text length, it works for a variety of document sizes. Furthermore, when determining the frequency with which a term appears in a particular text, TF (Term Frequency) accounts for changes in document lengths. On the other hand, IDF compares a word's meaning throughout the corpus to common, meaningless phrases like "that," "of," and "is". We may eliminate extraneous terms and concentrate on the most pertinent ones thanks to this strategy.

V. Results

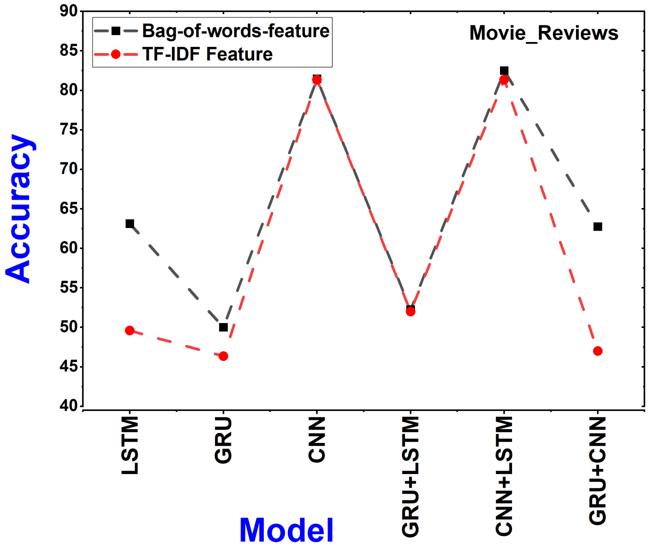


Fig. 8. Using six different DL algorithms with TF-IDF and Bag of Words, the accuracy score for movie reviews.

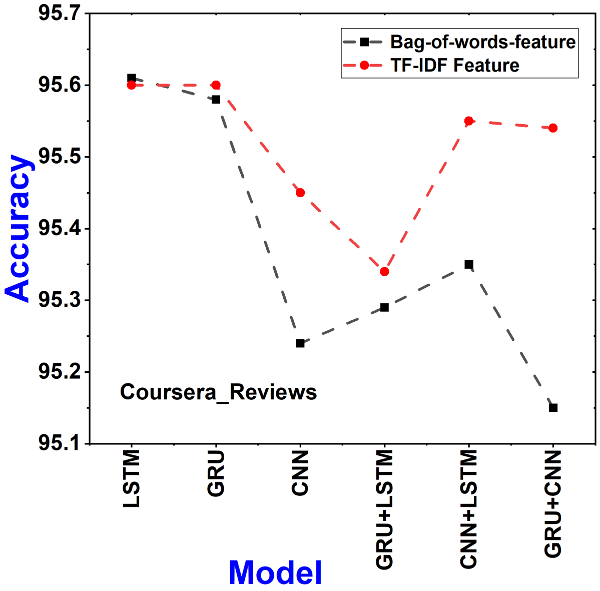


Fig. 9. Accuracy score for coursera reviews using six DL algorithms with TF-IDF and bag of words.

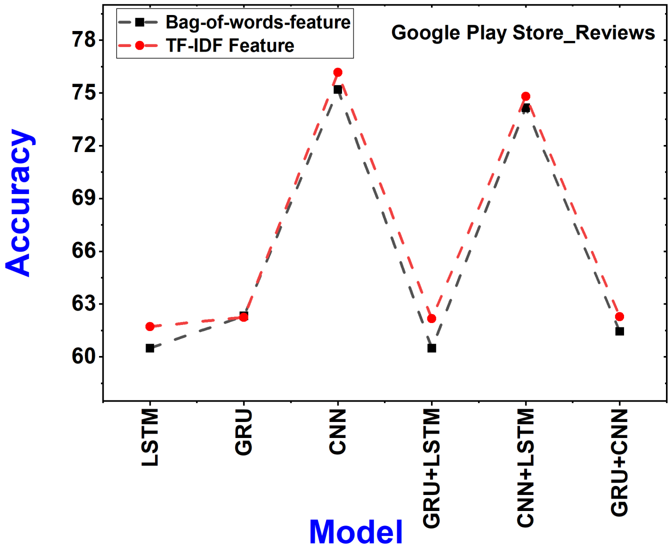


Fig. 10. Six DL algorithms with TF-IDF and bag of words were used to calculate the accuracy score for reviews on the google play store.

In this section, following the training of six distinct deep learning models (LSTM, GRU, CNN, LSTM+GRU, CNN+LSTM, and CNN+GRU), it became evident that the CNN+LSTM model consistently achieved higher accuracy scores across all the datasets, i.e., Movie, Coursera, Google Play Store, and Flipkart reviews compared to other deep learning models. The outcomes are displayed in Fig. 8, 9, 10, 11. Accuracy utilizing deep learning methods for different data sets is shown in Table I. We also observed that when the data set contains more reviews, deep learning models are more accurate in predicting compared to machine learning models. While CNNs on their own are typically not anticipated to surpass LSTM or GRU in terms of sentiment analysis performance due to their limited capacity to capture sequential dependencies, the combination of CNNs with RNNs, like CNN+LSTM often proves to be more effective than standalone models, shown in the Fig. 8, 9, 10, 11. Because these models can leverage the strength of CNNs in capturing local patterns and combine it with the sequential modeling capabilities of LSTMs. The

TABLE I. Accuracy Utilizing Deep Learning Methods for Different Data Sets

Model	Features	Movie	Coursera	Google Play Store	Flipkart
LSTM	Bag-of-words-feature	63.12	95.61	60.49	89.38
	TF-IDF Feature	49.59	95.60	61.72	89.94
GRU	Bag-of-words-feature	50.00	95.58	62.33	87.35
	TF-IDF Feature	46.34	95.60	62.25	88.91
CNN	Bag-of-words-feature	81.46	95.24	75.20	92.18
	TF-IDF Feature	81.34	95.45	76.16	93.08
GRU + LSTM	Bag-of-words-feature	52.25	94.29	60.49	82.42
	TF-IDF Feature	52.0	95.34	62.17	87.88
CNN + LSTM	Bag-of-words-feature	82.50	95.35	74.16	92.25
	TF-IDF Feature	81.36	95.55	74.80	93.45
GRU + CNN	Bag-of-words-feature	62.75	95.15	61.45	89.38
	TF-IDF Feature	47.00	95.54	62.28	89.97

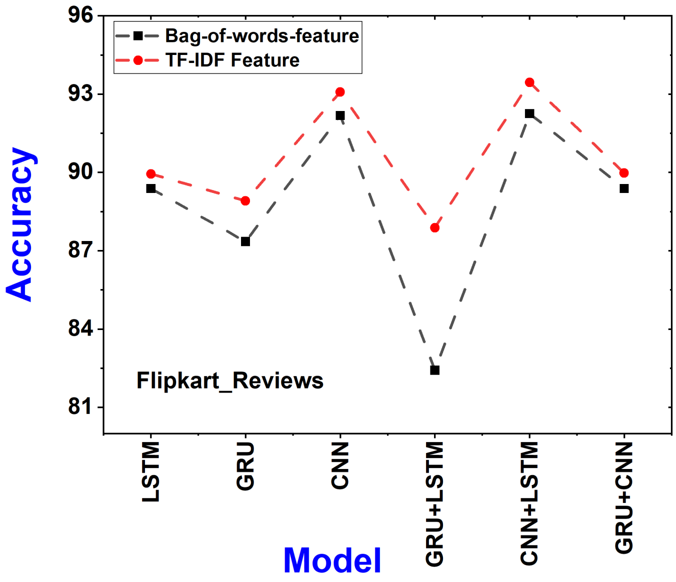


Fig. 11. The accuracy score for flipkart reviews is based on six DL algorithms, including bag of words and TF-IDF.

CNN component is capable of extracting high-level textual features, whereas the LSTM excels at capturing long-term dependencies. We also observed that the Combination of CNN with LSTM performs better than the combination of CNN with GRU. Also, we observed the combination of GRU+ LSTM will not lead good results compared to other deep learning models. This is due to a combination of both GRU and LSTM, which are similar in their functioning, and combining them does not introduce a significantly new perspective for sentiment analysis. Also, we can say that when CNNs are used for text-based tasks, they are typically employed in combination with RNNs or as part of more complex architectures, such as the Attention-based models, Transformers, or BERT-based models. These designs have performed admirably in a variety of NLP tasks, including sentiment analysis, by capturing both local and global patterns in text data. Ultimately, it was observed that Deep Learning models (LSTM, GRU, CNN, LSTM+GRU, CNN+LSTM, and CNN+GRU) outperform machine learning techniques (Logistic Regression, KNN Classifier, Bernoulli Naive Bayes, Multinomial Naive Bayes, XGBoost, Decision Tree, Random Forest Classifier) in the prediction of sentiment analysis. We also notice that these models yield improved results when utilizing the TF-IDF feature extraction technique.

In addition to the previously mentioned findings, it is noteworthy to highlight the quantitative performance of the proposed CNN+LSTM and CNN+GRU models in comparison to other state-of-the-art deep learning algorithms across diverse datasets, namely Movie, Coursera, Google Play Store, and Flipkart. The assessment was conducted using both bag-of-words and TF-IDF features to comprehensively evaluate the models' capabilities.

The CNN+LSTM model, leveraging bag-of-words features, exhibited commendable accuracy scores of 82.50, 95.35, 74.16, and 92.25 across the respective datasets. Employing TF-IDF features, the CNN+LSTM model demonstrated consistently high accuracy scores of 81.36, 95.55, 74.80, and 93.45, except in the Movie review dataset, where TF-IDF slightly outperformed bag-of-words. Notably, in the Movie review dataset, bag-of-words not only outperformed TF-IDF for the CNN+LSTM model but also surpassed the other five algorithms in accuracy.

In stark contrast, the GRU+LSTM model, when utilizing bag-of-words features, showed lower accuracy scores of 52.25, 94.29, 60.49, and 82.42 across the Movie, Coursera, Google Play Store, and Flipkart datasets, respectively. Similarly, with TF-IDF features, the GRU+LSTM model exhibited accuracy scores of 52.0, 95.34, 62.17, and 87.88. It is noteworthy that, compared to the other 5 DL models, the performance of the GRU+LSTM model is relatively poor.

VI. Conclusions

This article compares and contrasts sentimental categorization and analysis using several Deep-learning techniques that are currently available. Finally, several significant insights are shown by Table I of model accuracies based on different features and datasets. TF-IDF features are superior to Bag-of-Words features in capturing the nuances of the text data, as demonstrated by the consistently superior performance of the models that used them. The GRU with TF-IDF features and the CNN with TF-IDF features demonstrated the highest accuracies across the separate models across several datasets, demonstrating their efficacy in sentiment analysis tasks. In the realm of sentiment analysis, deep learning models exhibit superior performance when contrasted with traditional machine learning models across various domains, including reviews on Coursera, Flipkart, movies, and the Google Play Store. It's important to remember, though, that the model's performance varied greatly based on the dataset that was employed.

The majority of the models in the data from the Coursera and Google Play Store had consistently excellent accuracy scores, whereas the Movie dataset had lower accuracy ratings. Additionally, Flipkart data showed that different models performed differently, with some doing remarkably well and others having difficulty. The performance of the combination models, like CNN + LSTM and GRU + CNN, was frequently good, demonstrating the advantages of integrating several neural network architectures for improved sentiment analysis. In actuality, the model and feature representation used should be based on the needs of the particular dataset and application. Even if TF-IDF features work well most of the time, they might not be the ideal option every time. Deep learning excels in sentiment analysis by automatically learning intricate patterns, capturing contextual dependencies, and enabling end-to-end feature extraction. Hierarchical feature learning and adaptability to diverse data types make deep learning models like CNNs and LSTMs effective, the unique properties of the data should be taken into account while deciding between combination models and individual models. Future work should explore enhancing the interpretability of deep learning models in sentiment analysis, addressing its nature through methods like attention mechanisms. Investigating efficient training strategies, and data-efficient approaches, and applying novel architectures, including transformer-based models, will contribute to advancing the field and making deep learning more accessible for sentiment analysis tasks.

References

[1] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[2] B. Liu *et al.*, "Sentiment analysis and subjectivity." *Handbook of natural language processing*, vol. 2, no. 2010, pp. 627–666, 2010.

[3] A. Tripathy, "Sentiment analysis using machine learning techniques," Ph.D. dissertation, 2017.

[4] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, 2022.

[5] A. Abbasi and H. Chen, "Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace," *ACM Trans. Inf. Syst.*, vol. 26, no. 2, apr 2008. [Online]. Available: https://doi.org/10.1145/1344411.1344413

[6] M. N. Sadiku, T. J. Ashaolu, A. Ajayi-Majebi, and S. M. Musa, "Artificial intelligence in social media," *International Journal of Scientific Advances*, vol. 2, no. 1, pp. 15–20, 2021.

[7] T. Balaji, C. S. R. Annavarapu, and A. Bablani, "Machine learning algorithms for social media analysis: A survey," *Computer Science Review*, vol. 40, p. 100395, 2021.

[8] S. Ramaswamy and N. DeClerck, "Customer perception analysis using deep learning and nlp," *Procedia Computer Science*, vol. 140, pp. 170–178, 2018.

[9] A. Iqbal, R. Amin, J. Iqbal, R. Alroobaea, A. Binmahfoudh, and M. Hussain, "Sentiment analysis of consumer reviews using deep learning," *Sustainability*, vol. 14, no. 17, p. 10844, 2022.

[10] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," *Image and Vision Computing*, vol. 65, pp. 3–14, 2017, multimodal Sentiment Analysis and Mining in the Wild Image and Vision Computing. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885617301191

[11] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," *Information Processing & Management*, vol. 58, no. 5, p. 102656, 2021.

[12] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for e-commerce product reviews in chinese based on sentiment lexicon and deep learning," *IEEE access*, vol. 8, pp. 23 522–23 530, 2020.

[13] M. E. Alzahrani, T. H. Aldhyani, S. N. Alsubari, M. M. Althobaiti, and A. Fahad, "Developing an intelligent system with deep learning algorithms for sentiment analysis of e-commerce product reviews," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[14] C. Udomcharoenchaikit, P. Boonkwan, and P. Vateekul, "Adversarial evaluation of robust neural sequential tagging methods for thai language," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 19, no. 4, pp. 1–25, 2020.

[15] H. Akmal, F. Hardalaç, and K. Ayturan, "A fetal well-being diagnostic method based on cardiotocographic morphological pattern utilizing autoencoder and recursive feature elimination," *Diagnostics*, vol. 13, no. 11, p. 1931, 2023.

[16] C. Singh, T. Imam, S. Wibowo, and S. Grandhi, "A deep learning approach for sentiment analysis of covid-19 reviews," *Applied Sciences*, vol. 12, no. 8, p. 3709, 2022.

[17] H. Kaur, S. U. Ahsaan, B. Alankar, and V. Chang, "A proposed sentiment analysis deep learning algorithm for analyzing covid-19 tweets," *Information Systems Frontiers*, pp. 1–13, 2021.

[18] M. S. Akhtar, T. Garg, and A. Ekbal, "Multi-task learning for aspect term extraction and aspect sentiment classification," *Neurocomputing*, vol. 398, pp. 247–256, 2020.

[19] D.-H. Pham and A.-C. Le, "Learning multiple layers of knowledge representation for aspect based sentiment analysis," *Data & Knowledge Engineering*, vol. 114, pp. 26–39, 2018.

[20] A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *Ieee Access*, vol. 6, pp. 13 949–13 957, 2018.

[21] M. K. Enduri, A. R. Sangi, S. Anamalamudi, R. C. B. Manikanta, K. Y. Reddy, P. L. Yeswanth, S. K. S. Reddy, and A. Karthikeya, "Comparative study on sentimental analysis using machine learning techniques," *Mehran University Research Journal Of Engineering & Technology*, vol. 42, no. 1, pp. 207–215, 2023.

[22] M. S. Divate, "Sentiment analysis of marathi news using lstm," *International journal of Information technology*, vol. 13, no. 5, pp. 2069–2074, 2021.

[23] Q. Lu, X. Sun, Y. Long, Z. Gao, J. Feng, and T. Sun, "Sentiment analysis: Comprehensive reviews, recent advances, and open challenges," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[24] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath, "Sentiment analysis using gated recurrent neural networks," *SN Computer Science*, vol. 1, pp. 1–13, 2020.

[25] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, "Sentiment analysis using convolutional neural network," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2015, pp. 2359–2364.

[26] R. Ni and H. Cao, "Sentiment analysis based on glove and lstm-gru," in *2020 39th Chinese Control Conference (CCC)*, 2020, pp. 7492–7497.

[27] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, "A hybrid cnn-lstm model for improving accuracy of movie reviews sentiment analysis," *Multimedia Tools and Applications*, vol. 78, pp. 26 597–26 613, 2019.

[28] S. Elzeheiry, W. A. Gab-Allah, N. Mekky, and M. Elmogy, "Sentiment analysis for e-commerce product reviews: Current trends and future directions," 2023.

Predictive Modeling of Landslide Susceptibility in Soft Soil Canal Regions: A Focus on Early Warning Systems

Dang Tram Anh, Luong Vinh Quoc Danh, Chi-Ngon Nguyen*
Can Tho University, Can Tho City, Vietnam

***Abstract*—The Mekong Delta (MD) has suffered significant losses in land resources, economic damage, and human and property casualties due to recent landslides. An early warning system for landslides is a valuable tool for identifying the effectiveness and timely detection of changes in the soil to promptly determine solutions and minimize damage caused by landslides in an area. In this study, we apply a machine learning approach based on the Long Short-Term Memory (LSTM) algorithm to experiment with early warning of landslide events on soft soil in the MD. Horizontal pressure, the change in inclination angles of the sensor pile due to the soil mass sliding in both the x and y directions, and the warning levels are determined based on the deformation and displacement of the soil along the riverbank, considered candidate factors for inputs in the model. Data from the established sensor system is used to train the model, creating a training and testing dataset of 374,415 samples. The accuracy of the detection and classification threshold of the system is proposed to be measured using the average F1 score derived from precision and recall values. The optimal prediction results are gleaned from an observational window of 4 minutes and 30 seconds to project roughly 2 hours into the future. The validation process resulted in recall, precision, and F1-score stands at 0.8232 with a remarkably low standard deviation of about 1%. The successful application of this research can help identify abnormal events leading to riverbank landslides due to loading, thereby creating conditions for developing a reliable information system to provide managers with the ability to suggest timely solutions to protect the lives, property of residents and infrastructures.**

Keywords*—Landslide early warning; soft soil; Mekong Delta; long short-term memory

I. Introduction

Landslides represent a formidable global challenge, exerting a profound toll on economies, depleting natural resources, and tragically affecting human lives [1], [2]. Despite their localized occurrence, the ramifications of landslides extend far beyond their immediate vicinity, resulting in significant devastation to vital infrastructure elements such as roads, bridges, and power lines. This, in turn, leads to the distressing loss of land, homes, and, most tragically, human lives [3]. Vietnam in Southeast Asia bears witness to a history marked by a succession of natural disasters, including but not limited to floods, the encroaching threat of rising sea levels, shifts in climate patterns, coastal erosion, and the peril of landslides. Among the vulnerable regions, the Mekong Delta stands out, situated in the southern expanse of the country, densely populated yet perilously exposed to the caprices of nature. Its topographical identity is characterized by extensive low-lying stretches, intricately interlaced with an intricate network of rivers, canals, and verdant wetland areas, rendering it particularly susceptible to the forces of erosion and inundation.

The Mekong Delta region experiences a notable uptick in landslide occurrences, a trend that tends to peak during the rainy months. This phenomenon is intricately linked to the seasonal patterns of precipitation and the unique geological characteristics of the area. As moisture-laden rains saturate the soil and increase the weight and pressure on slopes, the propensity for landslides escalates. This heightened vulnerability underscores the need for vigilant monitoring and proactive mitigation measures to safeguard the natural landscape and its communities. According to the statistical data provided by the National Steering Committee for National Disaster Prevention and Control - Vietnam [4], until September 2023, there have been 558 locations of riverbank landslide within the area, resulting in a total affected and lost land length of over 740 km. Among these are 81 hazardous landslide sites and 137 high-risk landslide sites, causing damage to over 200 houses, residents' properties, and infrastructure worth thousands of billions of Vietnamese dong (VND).

Throughout the initial nine months of 2023, Can Tho City, situated among the thirteen provinces in the Mekong Delta region, bore the brunt of an alarming surge in landslides. This period witnessed an unsettling total of over 30 landslide incidents. These events led to injuries sustained by two individuals, the complete submersion of eight houses into the river, partial collapses, and grave impacts on 19 other residences. The cumulative length of the riverbank affected by these events totaled 1976 meters. Furthermore, the region has grappled with an escalating frequency of landslides in recent years, a trend partly attributed to the transformative impact of human activities and urban development encroaching upon riverine areas, thereby altering the natural landscape [5]. This surge in landslide occurrences has had far-reaching consequences, significantly and adversely affecting the lives of local residents and impeding the region's broader socio-economic development.

A preliminary assessment by domestic experts on the causes of erosion and instability of riverbanks and coastlines in the Mekong Delta region indicates that erosion at the base of the slope, saturated slopes after prolonged heavy rains or floods, and variations in groundwater levels are the primary factors contributing to slope instability along the riverbank [6], [7]. In addition to these long-term causes, the load of the structure, particularly the dynamic load induced by traffic,

*Corresponding authors.

can act as triggering factors for landslides [8], [9]. The complex relationships between landslide disasters and the factors that trigger them remain unclear. This complexity makes it challenging to analyze these mechanisms of landslides using simple algebraic equations [10]. Hence, predicting the ground's stress behavior and the soil mass's displacement along the riverbank becomes crucial and challenging.

The exploration of landslide risk assessment commenced in the 1970s [11]. Since then, propelled by advancements in modern statistical science and the advent of machine learning techniques, deep learning has become a vital tool in landslide research, particularly for landslide detection [12], [13], [14], [15]. While conventional approaches often treat landslide displacement prediction as a static regression problem, it is imperative to acknowledge that landslides are dynamic systems. The influencing factors and displacement conditions at one moment profoundly influence the subsequent moment. Thus, an effective landslide displacement prediction model should integrate susceptibility modeling, which considers the accumulation of variables over time.

The primary objective of this research is to present a predictive model for landslides utilizing advanced deep learning. The approach is rooted in the integration of input data, which encompasses variables such as soil pressure, incline in the x and y directions, alongside temporal factors. By harnessing machine learning algorithms, the primary aim of this work is to unravel the intricate interplay between these input parameters and their consequential impact on riverbank landslides. In the context of this study, a specialized form of recurrent neural network known as LSTM neural network was utilized to predict the progression of landslide events, extracting profound insights from time series data. This architectural framework excels in capturing complex patterns and temporal relationships in sequential data, providing favorable conditions for the development of a robust reduction framework. To facilitate the landslide prediction experiments, this study deployed a data collection system based on sensors. Through training on the gathered dataset, the model has demonstrated effective predicting capabilities for early warning of landslide phenomenon through training on the gathered dataset.

The underlying objective of the research is to study the performance of deep learning neural networks like LSTM to devise an early warning landslide system. The model is evaluated in term of accuracy on the dataset collected from the various experiments along the riverbank. In Section II we have discussed the related work. Section III explains the sources of collected dataset, the adopted methodology, the evaluation method and discusses the results in the study. Section IV is the conclusion of study.

II. Related Works

Recent machine learning and neural network breakthroughs have prompted extensive research, with CNNs being favored for their effectiveness in diverse applications. By utilizing neural networks, many researchers have developed landslide susceptibility maps and spatial modeling of landslides in highly prone mountainous areas using artificial neural networks [12], [16], [17], [18], [19], [20], [21]. Applying machine learning techniques along with GIS information [15], nonlinear spatial models have been created [22]. Neural Networks [23], [24], [25], [26], Boosted Regression Trees [27], [28], the Wavelet Transform model [29], [30] and Random Forest [31], [32], [33] are among the AI algorithms applied to tackle the intricate problems of landslide assessment. The Random Subspace Space Fuzzy Rule-based Classifier (RSSCE) is utilized to predict landslide occurrences based on the analysis of rainfall data triggered by heavy rainfalls in hilly areas [34], [35].

Research on landslide displacement prediction has indicated that utilizing deep learning on time series data, specifically the Long Short-Term Memory (LSTM) model, offers accurate forecasts [36]. LSTM's ability to capture historical information reduces the complexity of triggering factors, making it suitable for predicting landslide trends. Its performance in identifying these factors based on spatiotemporal sensor data demonstrates its superiority over other algorithms [37]. By extracting precise correlations between temporal and spatial data, LSTM has captured intricate time-dependent dependencies, outperforming traditional mechanical models [38]. The study in the Three Gorges Reservoir area, based on experimental mode decomposition and LSTM, has demonstrated superior accuracy compared to other static models in landslide displacement prediction [39]. Besides, [40] also proposes a novel coupled method using LSTM neural networks and support vector regression (SVR) algorithm, focusing on decomposing cumulative displacement into trend and periodic terms. The author introduces ensemble models based on SVR to optimize the combination of LSTM and SVR results, aiming for better accuracy in landslide predictions. The research on the mechanisms of damage and long short-term memory model for landslide prediction aims to address surface degradation phenomena, focusing on short-term impacts. The study emphasizes the effectiveness of LSTM in predicting landslide behavior and assessing damage based on joint distribution characteristics in building disaster prediction models [41]. These studies have applied the LSTM model based on various data such as rainfall, reservoir level, change in reservoir level, humidity, elevation, and displacement. The prediction results have observed that the prediction performance of the LSTM model is suitable and superior to static models like SVR, BP, SVM, ARMA, and even the dynamic Elman model.

Besides, combined models have been highly successful in fields such as flooding [42], [43] and drought [44], which has led researchers to explore ensemble modeling in landslide prediction as well. Risk analysis and forecast of landslide hazards using LSTM-RNN and DBA-LSTM models based on rainfall changes and water level variations have shown high accuracy [45], [46]. Researchers have used the original sequence of landslide displacement as input for combined machine learning models to predict the movement of landslides in steep slopes [47], [48], [49], [50]. The prediction of landslide movement in the major branches of the Three Gorges Dam area in China is carried out through combined models such as LSTM-TAR VMDstacked, LSTM-FC models, and LSTM models combined with Weighted Moving Average (WMA) using rainfall and reservoir water level data in each cycle with high accuracy [50], [51]. A prediction model for slope displacement based on the LSTM neural network and the Singular Spectrum Analysis (SSA) algorithm, using survey and geotechnical monitoring data, has significantly improved the model's performance in the dataset for predicting displacement within the next 24 hours

[52]. Landslides are likely to occur more frequently, along with climate change and the increasing surface loads caused by human activities.

In general, the conducted studies have been carried out in mountainous areas with a geological structure of loose rock. The primary factors causing landslide movements are mainly related to the permeability of the surface soil layer on the slope due to rainfall. In contrast to the riverbank landslides in the Mekong Delta, where the substrate is composed of alluvial deposits, young deltaic sediment, and riverbank soil with a predominantly geological structure of clayey soil, which is soft and weak. Additionally, factors contributing to riverbank landslides can arise from various causes: the inherently weak and saturated soil substrate with poor load-bearing capacity, subsurface erosion, changes in groundwater levels along the riverbank resulting in seepage flow, and human activities (such as construction and traffic), depicted in Fig. 2a. The studies on data collection systems based on modern sensor technologies or remote sensing techniques to generate databases for training landslide prediction models are presented in Table I and Table II. Thus, the need for early warning to assess the stability of slopes is becoming imperative [53]. Despite the extensive use of LSTM algorithms for landslide prediction in mountainous terrains with rocky geological structures, they have not been applied to predict riverbank landslides occurring in soft soil formations.

In this study, a new method to predict riverbank landslides using deep learning combined with data collected from sensors was proposed. To facilitate this, an Internet of Things (IoTs) - based sensor data collection and monitoring system have been implemented at the experiment sites along the riverbank within the Mekong Delta region. The collected data is used for training the LSTM machine learning algorithm to predict the possibility of riverbank slope instability for the purpose of early warning of riverbank landslides.

III. Methods

A. System Architecture for Predicting Landslides

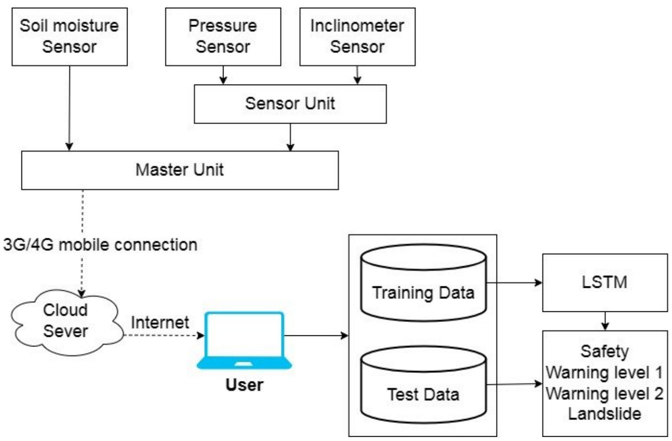


Fig. 1. System architecture for predicting landslides.

The architecture of the landslide prediction system is presented in Fig. 1. Our endeavor revolves around establishing a cutting-edge data collection framework focused on a diverse array of sensor devices. These instruments encompass a range of measurements, from monitoring soil moisture and inclination to gauging soil pressure. Central to this intricate setup is the Master Unit, a pivotal component acting as the central hub where these distinct sensors seamlessly converge to facilitate the precise acquisition of data. The Master Unit is a testament to seamless integration, meticulously designed to aggregate data from various sources with exceptional precision and efficiency.

In our commitment to ensure the accessibility and continuous availability of this data, we have dedicated substantial effort to engineer a robust data transmission mechanism. This designed system leverages the high-speed capabilities of 3G and 4G connectivity, enabling the collected data to be transferred from the Master Unit to our designated server. This data transmission process is characterized by its exceptional reliability and speed, forming an essential bridge that connects the physical realm of sensor data with the digital domain and ensures that this critical information is delivered promptly and with utmost accuracy.

The collected dataset undergoes a data preprocessing phase. This phase involves a blend of data cleansing, transformation, and normalization, all working in concert to ensure the data's integrity and suitability for subsequent analysis. A grid search method is employed on the training dataset to determine the appropriate number of layers of the LSTM method. The training dataset is split into two parts that comprise 80% and 20% of the dataset. The first part is used for training the model, and the second part is used to test the model's predictive ability. The crux of our methodology resides in utilizing deep learning architectures: LSTM. Our predictive framework relies on this algorithm model, which is predictable in time series within sequential data. The primary objective of this study is to forecast the warning level and landslide probability at time t, utilizing current time series data. This proactive approach significantly enhances the effectiveness of relocation efforts and minimizes potential damage in the aftermath of a landslide.

B. Experimental Setup for Data Collection System

In this work, to ensure consistency with the geological reality of the study, we excavated a channel (measuring $25m \times 3m$, with a slope ratio of $H = 0.5, V = 2$ on the banks of the Cai Sau channel in Can Tho city and tested the proposed system (Fig. 3). The geology at the research location is determined to be soft soil with the following physico-mechanical parameters: soil cohesion $C = 8.1kPa$, internal friction angle $\phi = 316'$, bulk weight $w = 1.571g/cm^3$, void ratio $e = 1.808$ and its initial water contents was 67%. The sensor pile is fixed at a distance of $0.1m$ from the channel bank edge (Fig. 4). The sensor node is composed of one Master unit and a Sensor unit. The soil pressure transducer is fixed on the sensor pile and inserted at a depth of 0.2 to 0.4 meters from the ground surface. The inclination sensor is compact (5 to $8mm$ in width and $10mm$ in length), installed in a waterproof plastic tube, and fixed on the sensor pile. The inclination sensor and soil pressure transducer are connected to the Sensor unit. The other item to be measured is soil moisture using a soil moisture sensor. This type of sensor is easy to use, plugged into the soil at a depth of $0.2m$ to observe and maintain the moisture of the canal bank slope in a saturated state throughout the

TABLE I. Monitoring Systems Provide Early Warning of Landslides

Landslide warning system	System	Input data	Topographic characteristics of the study area
In the world			
One sensor node is the collection of three types of sensors that are displacement sensor, pore pressure sensor and moisture sensor [54].	Landslide early warning system (LEWS)	Rain-fall, pore pressure, moisture and displacement	Mountainous regions
The Infrastructure Node (IN), the Subsurface Node (SN) and the Low-Cost Chain Inclinometer (LCI) [55].	LEWS based on IoTs technologies such as micro-electro-mechanical systems (MEMS) sensors and the LoRa (Long Range) communication protocol.	Subsurface deformation and ground seepage-water levels	Mountainous regions
Tilt sensors and volumetric water content sensors [56].	LEWS – MEMS	The tilting angles and moisture content	Mountainous regions
The on- site monitoring and collection nodes are composed of one STM32L071RBT6 microprocessor and one STM32 microprocessor processes the sensor data acquired [57].	Real-Time Monitoring System of Landslide - LoRa.	Rainfall, displacement, tilt and acceleration	Mountainous regions
Pressure sensors and strain gauges measure water pressure and soil displacement respectively [58].	Landslide Early Warning Wireless Sensor Networks (LEWS - WSNs) Ultra-wideband (UWB)	Soil moisture, water level, soil inclination and temperature	Mountainous regions
Volumetric water content (VWC) and pore water pressure (PWP) [59].	Local landslide early warning system (Lo-LEWS)	The calculated safety factor (FS), the temperature, the precipitation, the VWC and PWP monitored were used as input dataset for a supervised machine learning algorithm	Mountainous regions
In Viet Nam			
The proposed system consists of six sensor nodes and one rainfall station [60].	Rainfall-induced landslide early warning (EWMRIL) - ZigBee	Soil moisture, PWP, movement status, and rainfall.	Mountainous regions

testing process. Then, the Sensor unit and Soil moisture sensor are connected to the Master unit. Measurement data will be transmitted to the Cloud server via 3G/4G mobile networks by the Master unit every 60 seconds. Solar power is used to provide energy for the sensor node.

The applied load is in the form of a strip load, simulating the load of the structure along the riverbank. The pressure exerted on the canal bank is created by sandbags placed on a steel plate with an area of 0.6m x 1.4m, arranged around the sensor pile with a distance of $0.2m$ as shown in Fig. 3. The canal bank in the experimental area was saturated with artificial continuous rainfall of 15mm/h. After filling the canal with water, the canal's water level was changed in combination with waves generated (by wave generators) to create a scour hole at the base of the slope as in Fig. 2. The strip load was gradually increased until the landslide occurred. Each load level was 3.5 kPa. Data on soil pressure, pile tilt angle, and soil moisture were recorded during the experiment.

Measurement data of inclination angle and soil pressure during the destructive deformation of the ground at the pile position are shown in Fig. 5, respectively. It can be observed that the landslide occurs in a relatively short time, only a few minutes since the mechanism of landslide is sudden deformation. Values of the inclination angle in the two directions, x y, differ in magnitude and time because the sensor pile moves along with the sliding ground mass.

C. Dataset

We performed testing at 45 sites. Each experiment lasted approximately seven days. The loading and data recording started 24 hours after embedding the sensor piles into the riverbank soil mass, with the load level increasing from 3.0 kPa to 4.5 kPa at each level. Each loading level was applied for continuous 11-hour intervals. An illustrative example of a measured data set is the horizontal soil pressure recorded during the second loading phase, which was 1.16 kPa. The X and Y tilt angles showed minor variations in the first 3 hours due to soil structure consolidating pressure, increasing soil compactness at the measurement site. Loading gradually increased at specified intervals until signs of soil failure became apparent. Such data recorded at 12:35:58 on December 19, 2022, as shown in Fig. 5, indicated a horizontal soil pressure of 1.9 kPa under an applied total load of 13.5 kPa. At 22:13:28 on

TABLE II. Research on Landslides in the Mekong Delta

Research	Method	Results
Initial assessment on the causes of riverbank instability in Chau Thanh district, Hau Giang province [6].	Google Earth remote sensing images from 2006 to 2019 were used to assess the current status of riverside construction and erosion. The Analytic Hierarchy Process (AHP) was used to determine the impact levels of factors that cause riverbank instability. Using the AHP method and field survey can be extended to other areas in the Mekong Delta to analyze riverbank stability.	The survey and analysis results show that geology is the most affecting factor among the factors, and in combination with the encroaching construction of riversides to protection buffer areas, it creates the surcharge load reducing the stability coefficient of the riverbank. Besides, the curvature and flow velocity are also the causes of riverbed erosion and deformation, leading to an increase in the stiff slope which affects riverbank stability.
Analysis of factors affecting riverbank stability: case study at Cha Va river section, Vinh Long province [61].	Satellite images were loaded from Google Earth to analyze the current erosion of the river banks and the urbanization process along the river banks.	The integrated effects of soft soil, water level fluctuation, wave load, and surcharge loads is found to be the cause of instability of the Chavas river bank.
Assessment of the situation of landslides and sedimentation in the coastal area of Ca Mau and Bac Lieu province from 1995-2010 using remote sensing and GIS technology [62].	Research using remote sensing images Landsat	Assess shoreline changes erosion or deposition processes
Monitoring developments in Cu Lao Dung's coastline using remote sensing image analysis technology [63].	Multispectral Landsat images were utilizad for the analysis	Analyze erosion and deposition locations with specific values
Monitoring erosion and accretion situation in the coastal zone at Kien Giang province [64].	The study applied Normalized Difference Water Index (MNWI) method and water level extraction using LANDSAT imagery from 1975 to 2015 for highlight the shoreline.	Analysis was identified erosion and accretion areas based on shoreline changes and land use influenced by landslides and deposition

the same day, the horizontal soil pressure measured was 2.11 kPa and showed a decreasing trend. Subsequently, continuous changes in soil pressure and sensor pile tilt were observed, accompanied by the appearance of cracks on the canal bank, as shown in Fig. 4 (small cracks visible). This indicated that the soil had reached its ultimate limit state. By 23:34:58 on December 19, 2022, when the horizontal soil pressure reached 2.13 kPa, an additional loading level was applied, raising the total load to 18 kPa. The horizontal soil pressure and tilt angle remained relatively stable for about 1 hour and 30 minutes, while monitoring revealed the gradual expansion of cracks. Simultaneously, the soil pressure data showed a rapid decrease, and the tilt angle of the sensor pile experienced a sudden significant change. Finally, the canal bank slope completely slid after the horizontal soil pressure reached its peak value of 2.5 kPa. Data measured from sensors will be saved as data files on the memory card. Then, the Master device will scan the memory card and transmit the data files to the Cloud server. Users can download the measured data for analysis and real-time observation through the Web server. First, the dataset "landslide monitoring.csv" was downloaded. The monitoring dataset includes date and time, soil pressure, soil moisture, and inclination angles due to soil movement and landslides. Data was collected at one-second intervals throughout the loading experiment until the landslide occurred. We can use this data to address the prediction issue for the next two hours based on changes in ground pressure and soil displacement in the preceding hours. Next, the data was labeled and categorized according to warning levels. Then, the data was transformed into a supervised learning problem. We normalized the data for model training using a 30-second sliding window approach. After reprocessing, we obtained 374,415 samples, split into a training set comprising 70% (262,090 samples) and a test set comprising 30% (112,325 samples). The soil moisture content in the experimental conditions was always saturated; thus, this value was not included in the input dataset. There were three input variables: ground pressure (p), X-axis inclination angle (x angle), Y-axis inclination angle (y angle), warning level, and warning label arranged in 5 columns in the input data table.

D. Warning Thresholds

The calculated ultimate bearing capacity of the soil foundation using the finite element method based on geological parameters at the experimental site was determined to be 20.61 kPa. The analysis results of the finite element model under external loading show: (1) when the soil stress is below 40 percent of the ultimate bearing capacity, the soil remains in a safe state; (2) when the soil stress reaches a ratio below 70 percent of the ultimate bearing capacity, the soil starts deforming and consolidating; (3) similarly, when the ratio increases to 80 percent, the soil shows significant deformation and displacement; (4) finally, when this ratio exceeds 80

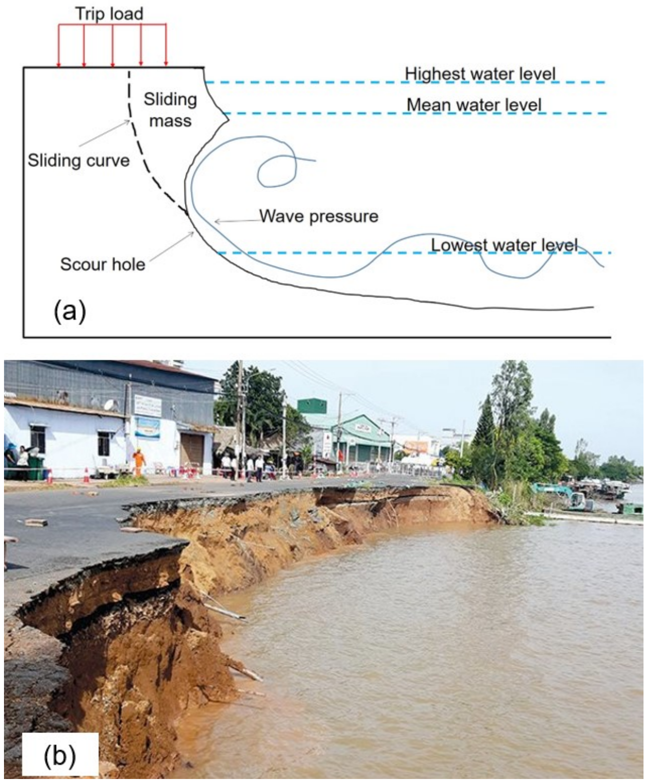


Fig. 2. Riverbank landslide: (a) Diagram of the main factors contributing to slope deformation; (b) Landslide submerges a section of national highway 91 into the hau river in an Giang province, Vietnam [65].

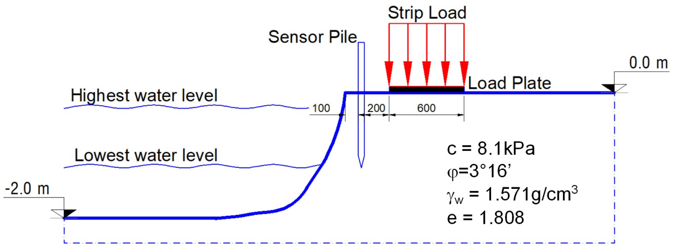


Fig. 3. Dimensions of the experimental canal.

percent, the analysis results indicate complete deformation and destruction of the soil. The analysis results using the numerical method extracted from Plaxis software are presented in Fig. 6, 7, 8, 9 with the surveyed load levels of 30%, 63%, 82.5% and 99.5% of the ultimate bearing capacity, respectively. Fig. 9 clearly shows the sliding curve of the riverbank soil mass (highlighted in orange) and the failure strain area (highlighted in red). The results of soil pressure values in the horizontal direction and displacement obtained from the measurement system showed similarities in pressure ratios with the model analysis results. Therefore, we have proposed pressure ratio values to determine the warning thresholds as shown in Table III below, where q_{ult} is the ultimate bearing capacity of the ground.

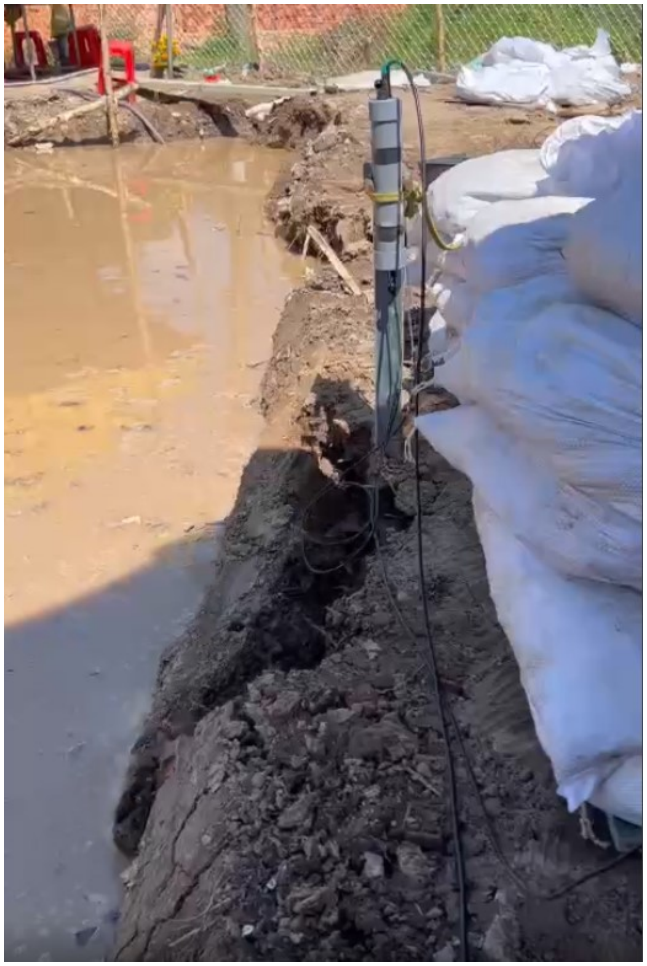


Fig. 4. Photo taken at the experimental site.

TABLE III. Pressure Ratio Values to Determine the Warning Thresholds

Soil pressure factor	Displacement	Level of warning
k $\leq$ 40%.qult	-	Safety
40%.qult <k $\leq$ 70%.qult	Narrow oscillation angle	Warning level 1
70%.qult <k $\leq$ 80%.qult	The oscillation angle gradually widens	Warning level 2
k >80%.qult	Sudden fluctuation in the oscillation angle	Landslide

E. Long Short-Term Memory

The architecture of an LSTM neural network is explicitly designed to handle sequential data, making it a powerful tool for tasks like time series prediction, natural language processing, and more. At its core, an LSTM network comprises individual LSTM cells that work together in a chain-like structure (see Fig. 11). Each LSTM cell (see Fig. 10) contains three main gates: the Forget Gate, the Input Gate, and the Output Gate, along with a Candidate Gate. These gates, implemented using a combination of activation functions and weights, control the flow of information through the cell. The Forget Gate determines what information to discard from the previous cell state. The Input Gate decides what new information to store in the cell state, and the Output Gate regulates what information to expose to the output. This ar-

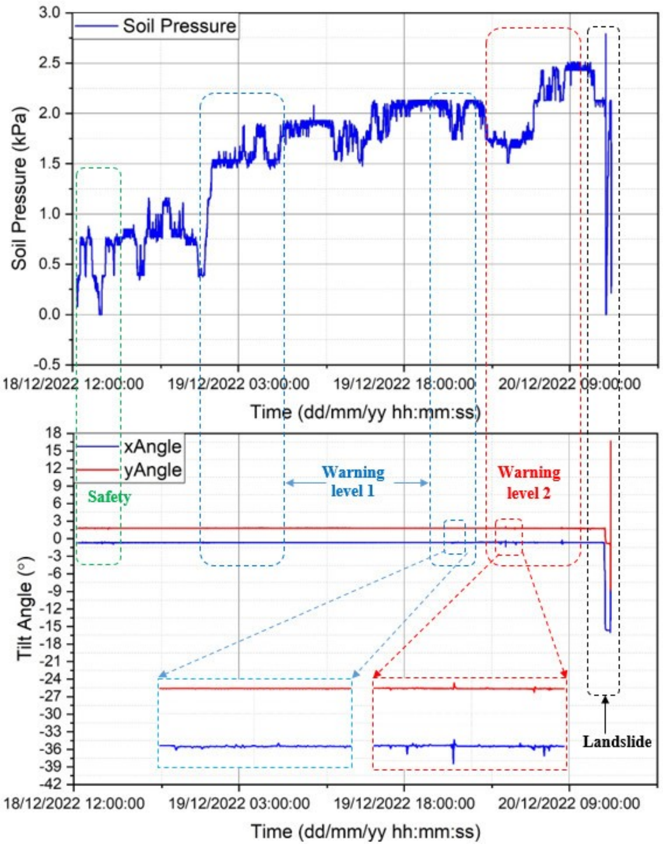


Fig. 5. A specific illustration of warning thresholds based on soil pressure and inclination displacement data of sensor piles.

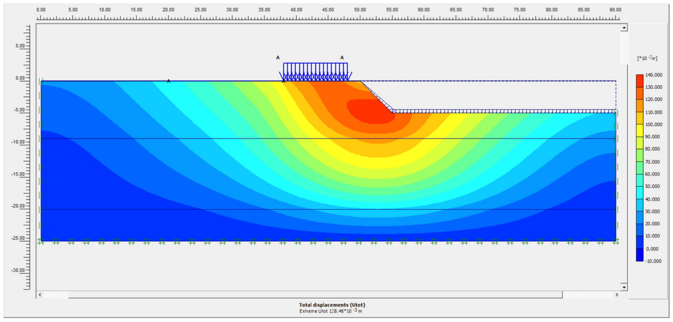


Fig. 6. The total displacement result at the load level q=6kPa.

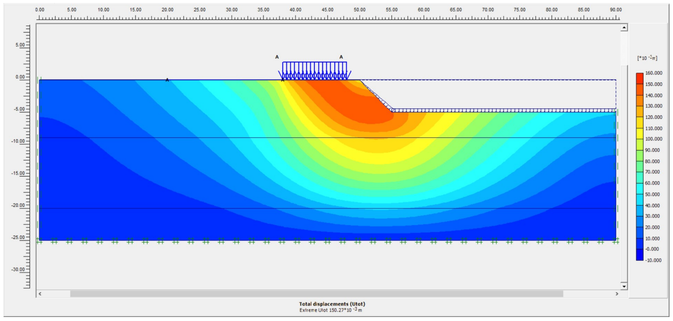


Fig. 7. The total displacement result at the load level q=13kPa.

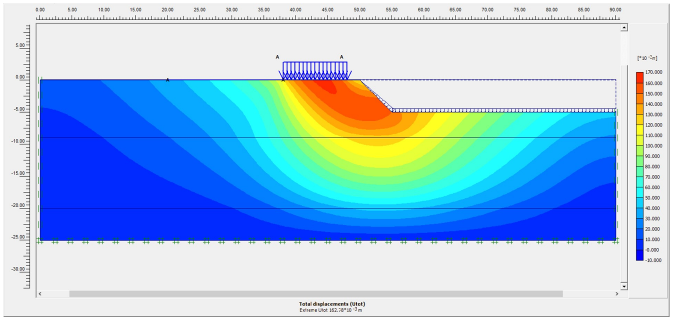


Fig. 8. The total displacement result at the load level q=17kPa.

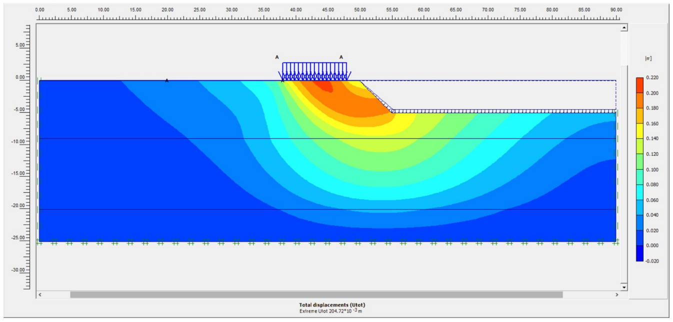


Fig. 9. The total displacement result at the load level q=20.5kPa.

chitectural design enables LSTMs to effectively capture long-term dependencies in sequential data, addressing the issue of vanishing or exploding gradients often encountered in standard RNNs.

The approach with LSTM models for landslide prediction represents a significant stride in harnessing advanced technology for geohazard mitigation. LSTM, a specialized form of recurrent neural network (RNN), excels in handling sequential data, making it particularly well-suited for predicting time-dependent variables associated with landslides. This is paramount in regions prone to geological instability, where timely warnings are crucial for safeguarding lives and infrastructure. One critical strength of LSTM models is their ability to capture intricate temporal dependencies and patterns within time series data. By incorporating memory cells that can retain and update information over time, LSTMs excel in modeling sequences characterized by long-term dependencies. This dynamic memory retention mechanism empowers the model to discern subtle shifts in environmental factors leading to a potential landslide event. Consequently, LSTM-based predictive models stand at the forefront of cutting-edge geotechnical research, offering a promising avenue for enhancing early warning systems and ultimately minimizing the impact of landslides on vulnerable communities.

The equations for an LSTM cell are as follows. In these equations, x_t represents the input at time t, h_t represents the hidden state at time t, c_t represents the cell state at time t, and f_t, i_t, g_t, and o_t represent the forget gate, input gate, cell gate, and output gate respectively.

1) Forget Gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

2) Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

3) Candidate Cell State:

$$g_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

4) Update Cell State:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot g_t$$

5) Output Gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

6) Hidden State:

$$h_t = o_t \cdot \tanh(c_t)$$

Here,

- σ represents the sigmoid activation function.
- $\tanh$ represents the hyperbolic tangent activation function.
- W_f, W_i, W_c, and W_o are weight matrices specific to the forget gate, input gate, cell gate, and output gate, respectively.
- b_f, b_i, b_c, and b_o are bias vectors associated with the respective gates.

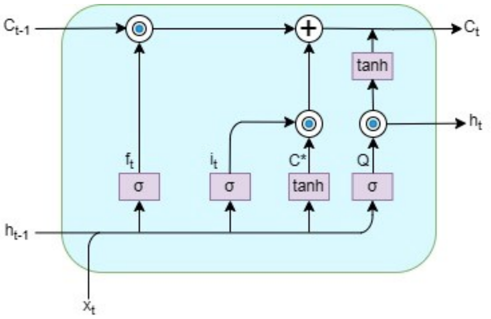


Fig. 10. The LSTM cell operation.

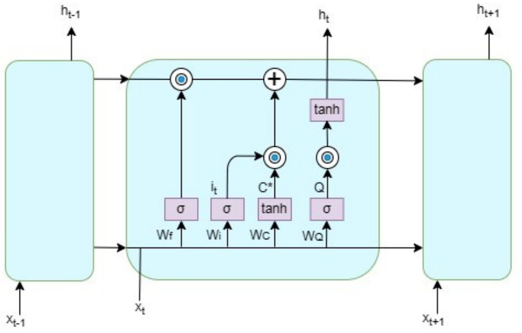


Fig. 11. Structure of LSTM.

F. Landslide Prediction with LSTM

LSTM networks resolve the vanishing gradient issue that hinders traditional RNN training by utilizing a memory cell capable of retaining information over extended durations. This quality makes them highly effective for modeling time series sequences, a crucial aspect in predicting landslides due to their complex temporal nature influenced by various factors.

The dataset gathered from 45 experimental points is vital for implementing the LSTM algorithm in predicting landslide warning levels. This extensive time series monitoring dataset captures a multitude of critical attributes, including ground pressure (p), the X-axis inclination angle (x angle), the Y-axis inclination angle (y angle), and the respective warning levels recorded at various experimental sites situated along the banks of Rach Cai Deep River. It's important to note that this dataset was meticulously collected at 30-second intervals, a frequency maintained throughout the loading test, persisting until a landslide event occurred. To formulate the predictive challenge effectively, we have stratified it into four levels: safety, warning level 1, warning level 2, and landslide. The very essence of this endeavor is rooted in the comprehensive time series data collected in the context of six scenarios. The particulars of these six distinct experimental scenarios are elucidated in depth in the experimental results section.

Furthermore, our experiments extend beyond mere data collection, encompassing the proactive prediction of landslide levels. This predictive process spans specific time intervals, specifically after 2 hours, 2.5 hours, 2.7 hours, 4.3 hours, 4.8 hours, and 5.5 hours post-initiation. This time horizon is of particular significance, representing a window during which strategic asset relocation can effectively mitigate the potential loss of valuable assets and infrastructure. For training and evaluating the LSTM model, the entire dataset is thoughtfully partitioned into two subsets. A substantial 70% of the data is designated for training, with the remaining 30% preserved for testing and validation. This partitioning strategy ensures the robustness and accuracy of the model's performance, affirming its capacity to effectively predict landslide warning levels under diverse temporal and environmental conditions.

G. Evaluation Method

In multiclass classification, the performance of a classification model is typically evaluated using various metrics to assess its accuracy and effectiveness in classifying data into multiple classes. The evaluation metrics used include accuracy, precision, and recall [66]. The F1 score is the harmonic mean of precision and recall, representing a balanced average between accuracy and recall [19], [67]. It ranges between 1 and 0. The highest possible value of F1 is 1.0, and an F1 score approaching 1 signifies perfect precision and recall, indicating high confidence and reliability of the algorithm in predicting landslide warning levels [68]. Detailing the indices presented in the documents [69], [70]. Common evaluation metrics include:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{1}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}} \tag{2}$$

$$\text{Recall (Sensitivity)} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \quad (3)$$

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

H. Experimental Results for Predicting Landslide

The LSTM model under study boasts a tri-layered structure, with each layer densely populated by 512 neurons. Despite its apparent complexity, the architecture remains streamlined and efficient, designed expressly for seamless operation on low-performance laptop hardware. The training of this model employs the AdamW optimizer, set at a learning rate of 0.0001. To ensure robustness and repeatability in our results, we conducted each experiment three times, then averaged the F1-scores, noting their standard deviation. Although the training was set for a potential maximum of 100 epochs, early stopping was incorporated with a patience of 8 epochs. Remarkably, most experiments converged to their optima around the 50 epochs. The model uses the CrossEntropyLoss function for optimization. Despite its depth and intricacy, the model remains relatively lightweight with 5.3 million trainable parameters, resulting in an estimated total model size of 21 MB. The forecast performance is depicted in Table IV.

TABLE IV. The Forecast Performance of our Model

Observed window	Forecast at	F1-score
30min	330min (∼5.5h)	0.8088 ± 0.0380
22min	292min (∼4.8h)	0.7688 ± 0.0136
22min	262min (∼4.3h)	0.7879 ± 0.0053
14min 30sec	164min 30sec (∼2.7h)	0.7927 ± 0.0319
04min 30sec	154min 30sec (∼2.5h)	0.7756 ± 0.0107
04min 30sec	124min 30sec (∼2h)	0.8232 ± 0.0122

In this research framework, the forecasting setup is delineated: The 'observed window' denotes the uninterrupted period during which data is observed. Data acquisition involved procuring three distinct time series, e.g., pressure, x-angle, and y-angle, from sensors strategically positioned on the experimental pole. The positioning strategy is squarely within the ambit of multivariate time-series forecasting [71], [72]. The forecasting objective seeks to predict events in the vicinity of the pole for a specified future time point. For example, if it monitors the three-time series for 4 minutes and 30 seconds, the predictive model aims to forecast events at the 124-minute and 30-second mark—roughly a 2-hour projection into the future.

I. Remark and Discussion

Fig. 4 illustrates the appearance of cracks on the ground, gradual soil mass horizontal displacement, and the tilted state of the sensor pile. The canal bank is progressively damaged, starting with the scour hole at the base of the slope. The canal bank is saturated after the rains and the increasing applied load. By monitoring the gradual expansion of cracks along the canal banks and around the sensor pile, the measured soil pressure data and observing the tilting behavior of the sensor pile before the soil mass ultimately failure took place over more than 12 hours. This is also identified as shown in Fig. 5 from the time the Level 1 warning signs appeared for the second time, then progressed to Level 2 warning until the soil mass slid into the canal, it took place from 21:00:00 on December 19, 2022, to 12:00:00 on December 20, 2022, equivalent to approximately 15 hours. Such behavior could be used as a signal for early warning. It should be noted that the pressure values and the behavior of tilting before failure vary case-by-case. Thus, the criteria for issuing warnings should be carefully determined.

From a research standpoint, while the F1-score demonstrates commendable performance at the 330-minute forecast time point, it is imperative to note a relatively higher standard deviation, approximately 3.8%. Our optimal prediction results are gleaned from an observational window of 4 minutes and 30 seconds to project roughly 2 hours into the future. For instance, the F1-score stands at 0.8232 with a notably low standard deviation of about 1%. However, by conducting experiments over various observational windows, we intend to establish a reliable database to provide in-depth insights for managers to implement appropriate solutions to protect the lives of residents and infrastructure.

IV. Conclusion

In this work, we have presented the establishment of an experimental model for monitoring the landslide of riverbank soil in the Mekong Delta region. The data collected from the experiment show good agreement with finite element method calculations. We propose warning thresholds based on soil pressure, deformation, and displacement data under saturated soil conditions. The results of this study contribute to the establishment of a system to predict the likelihood of extreme events at a specific time in the future based on 374,415 samples and three input parameters, including soil pressure, inclination of the sensor pile in directions towards and along the riverbank, warning level, and warning label, using the LSTM method. The experiment with an observation window of 4 minutes and 30 seconds indicates that the optimal result for predicting warning events is around 2 hours in the future. Through experimental results, the LSTM model effectively predicts early warning landslides for riverbanks with soft soil characteristics under external loading. The findings of this research can be applied in developing early warning systems for riverbank landslides in the Mekong Delta region.

References

[1] P. T. T. Ngo, M. Panahi, K. Khosravi, O. Ghorbanzadeh, N. Kariminejad, A. Cerda, and S. Lee, "Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of iran," *Geoscience Frontiers*, vol. 12, no. 2, pp. 505–519, 2021.

[2] M. Azarafza, M. Azarafza, H. Akgün, P. M. Atkinson, and R. Derakhshani, "Deep learning-based landslide susceptibility mapping," *Scientific reports*, vol. 11, no. 1, p. 24112, 2021.

[3] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. R. Meena, D. Tiede, and J. Aryal, "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sensing*, vol. 11, no. 2, p. 196, 2019.

[4] "Viet nam disaster and dyke ganagement authority. dong bang song cuu long: Sat lo bua vay," https://phongchongthientai.mard.gov.vn/Pages/dong-bang-song-cuu-long-sat-lo-bua-vay.aspx, (in Vietnamese).

[5] N. T. H. Diep, C. T. Nguyen, P. K. Diem, N. X. Hoang, and A. A. Kafy, "Assessment on controlling factors of urbanization possibility in a newly developing city of the vietnamese mekong delta using logistic regression analysis," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 126, p. 103065, 2022.

[6] L. T. Phat, D. V. Duy, C. T. Hieu, N. T. An, K. Lavane, and T. V. Ty, "Initial assessment on the causes of riverbank instability in chau thanh district, hau giang province," *Tap chi Khi tuong Thuy van*, vol. 740, pp. 57–73, 2022.

[7] L. X. Tu, T. B. Hoang, V. Q. Thanh, D. P. Wright, A. T. Hansan, and D. T. Anh, "Evaluation of coastal protection strategy and proposing multiple lines of defense under climate change of the mekong delta for sustainable shoreline protection," *Ocean and Coastal Management*, vol. 228, p. 106301, 2022.

[8] A. Xing, G. Wang, B. Li, Y. Jiang, Z. Feng, and T. Kamai, "Long-runout mechanism and landsliding behaviour of large catastrophic landslide triggered by heavy rainfall in guanling, guizhou, china," *Canadian Geotechnical Journal*, vol. 52, no. 7, pp. 971–981, 2015.

[9] J. Zhuang, J. Peng, G. Wang, I. Javed, Y. Wang, and W. Li, "Distribution and characteristics of landslide in loess plateau: A case study in shaanxi province," *Engineering Geology*, vol. 236, pp. 89–96, 2018.

[10] S. Xu and R. Niu, "Displacement prediction of baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in three gorges area, china," *Computers & geosciences*, vol. 111, pp. 87–96, 2018.

[11] P. Reichenbach, M. Rossi, B. Malamud, M. Mihir, and F. Guzzetti, "A review of statistically-based landslide susceptibility models," *Earth-science reviews*, vol. 180, pp. 60–91, 2018.

[12] E. Sevgen, S. Kocaman, H. A. Nefeslioglu, and C. Gokceoglu, "A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ann and random forest," *Sensors*, vol. 19, no. 18, p. 3940, 2019.

[13] M. I. Sameen, B. Pradhan, and S. Lee, "Application of convolutional neural networks featuring bayesian optimization for landslide susceptibility assessment," *Catena*, vol. 186, p. 104249, 2020.

[14] D. Sun, H. Wen, D. Wang, and J. Xu, "A random forest model of landslide susceptibility mapping based on hyperparameter optimization using bayes algorithm," *Geomorphology*, vol. 362, p. 107201, 2020.

[15] M. Conforti and F. Ietto, "Modeling shallow landslide susceptibility and assessment of the relative importance of predisposing factors, through a gis-based statistical analysis," *Geosciences*, vol. 11, no. 8, p. 333, 2021.

[16] Y. Wang, Z. Fang, and H. Hong, "Comparison of convolutional neural networks for landslide susceptibility mapping in yanshan county, china," *Science of the total environment*, vol. 666, pp. 975–993, 2019.

[17] L. Bragagnolo, R. da Silva, and J. Grzybowski, "Artificial neural network ensembles applied to the mapping of landslide susceptibility," *Catena*, vol. 184, p. 104240, 2020.

[18] C. Yu and J. Chen, "Landslide susceptibility mapping using the slope unit for southeastern helong city, jilin province, china: a comparison of ann and svm," *Symmetry*, vol. 12, no. 6, p. 1047, 2020.

[19] S. L. Ullo, A. Mohan, A. Sebastianelli, S. E. Ahamed, B. Kumar, R. Dwivedi, and G. R. Sinha, "A new mask r-cnn-based method for improved landslide detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3799–3810, 2021.

[20] T. Liu, T. Chen, R. Niu, and A. Plaza, "Landslide detection mapping employing cnn, resnet, and densenet in the three gorges reservoir, china," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11 417–11 428, 2021.

[21] W. L. Hakim, F. Rezaie, A. S. Nur, M. Panahi, K. Khosravi, C.-W. Lee, and S. Lee, "Convolutional neural network (cnn) with metaheuristic optimization algorithms for landslide susceptibility mapping in icheon, south korea," *Journal of environmental management*, vol. 305, p. 114367, 2022.

[22] M. Marjanović, B. Bajat, B. Abolmasov, and M. Kovaćević Marjanović, "Machine learning and landslide assessment in a gis environment," *GeoComputational Analysis and Modeling of Regional Systems*, vol. 19, no. 18, pp. 191–213, 2018.

[23] S. N. Selamat, N. A. Majid, and A. M. Taib, "A comparative assessment of sampling ratios using artificial neural network (ann) for landslide predictive model in langat river basin, selangor, malaysia," *Sustainability*, vol. 15, p. 861, 2023.

[24] H. Masruroh, A. S. Leksono, S. Kurniawan, and Soemarno, "Developing landslide susceptibility map using artificial neural network (ann) method for mitigation of land degradation," *Journal of Degraded and Mining Lands Management*, vol. 10, no. 3, pp. 4479–4494, 2023.

[25] F. Abbas, F. Zhang, F. Abbas, M. Ismail, J. Iqbal, D. Hussain, G. Khan, A. F. Alrefaei, and M. F. Albeshr, "Landslide susceptibility mapping: Analysis of different feature selection techniques with artificial neural network tuned by bayesian and metaheuristic algorithms," *Remote Sens*, vol. 15, p. 4330, 2023.

[26] M. Daviran, M. Shamekhi, R. Ghezelbash, and A. Maghsoudi, "Landslide susceptibility prediction using artificial neural networks, svms and random forest: hyperparameters tuning by genetic optimization algorithm," *International Journal of Environmental Science and Technology*, vol. 20, no. 1, pp. 259–276, 2023.

[27] S. Saha, A. Arabameri, A. Saha, T. Blaschke, P. T. T. Ngo, V. H. Nhu, and S. S. Band, "Prediction of landslide susceptibility in rudraprayag, india using novel ensemble of conditional probability and boosted regression tree-based on cross-validation method," *Science of the total environment*, vol. 764, p. 142928, 2021.

[28] Q. Zhu, A. Arabameri, M. Santosh, J. Egbueri, and J. Agbasi, "Integrated assessment of landslide susceptibility in the kalaleh basin, golestan province, iran using novel svr-goa ensemble validated with brt, ann, and elastic net models," *Environmental Science and Pollution Research*, 2023.

[29] G. Teza, S. Cola, L. Brezzi, and A. Galgaro, "Wadenow: A matlab toolbox for early forecasting of the velocity trend of a rainfall-triggered landslide by means of continuous wavelet transform and deep learning," *Geosciences*, vol. 12, p. 205, 2022.

[30] Y. Liu, G. Teza, L. Nava, Z. Chang, M. Shang, D. Xiong, and S. Cola1, "Deformation evaluation and displacement forecasting of baishuihe landslide after stabilization based on continuous wavelet transform and deep learning," *Under Review at Natural Hazards*, 2023.

[31] M. Krkac, S. B. Gazibara, Z. Arbanas, M. Secanj, and S. M. Arbanas, "A comparative study of random forests and multiple linear regression in the prediction of landslide velocity," *Landslides*, vol. 17, p. 2515–2531, 2020.

[32] V.-H. Dang, N.-D. Hoang, L.-M.-D. Nguyen, D. T. Bui, and P. Samui, "A novel gis-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility," *Forests*, vol. 11, no. 1, p. 118, 2020.

[33] D. Sun, H. Wen, D. Wang, and J. Xu, "A random forest model of landslide susceptibility mapping based on hyperparameter optimization using bayes algorithm," *Geomorphology*, vol. 362, p. 107201, 2020.

[34] B. Shi, T. Zeng, C. Tang, L. Zhang, Z. Xie, G. Lv, and Q. Wu, "Landslide risk assessment using granular fuzzy rule-based modeling: A case study on earthquake-triggered landslides," *IEEE Access*, vol. 9, pp. 135 790–135 802, 2021.

[35] S. Badola, V. N. Mishra, S. Parkash, and M. Pandey, "Rule-based fuzzy inference system for landslide susceptibility mapping along national highway 7 in garhwal himalayas, india," *Quaternary Science Advances*, vol. 11, p. 107201, 2023.

[36] A. Aggarwal, M. Alshehri, M. Kumar, O. Alfarraj, P. Sharma, and K. R. Pardasani, "Landslide data analysis using various time-series forecasting models," *Computers & Electrical Engineering*, vol. 88, p. 106858, 2020.

[37] L. Xiao, Y. Zhang, and G. Peng, "Landslide susceptibility assessment using integrated deep learning algorithm along the china-nepal highway," *Sensors*, vol. 18, no. 12, p. 4436, 2018.

[38] P. Xie, A. Zhou, and B. Chai, "The application of long short-term memory(lstm) method on displacement prediction of multifactor-induced landslides," *IEEE Access*, vol. 7, pp. 54 305–54 311, 2019.

[39] S. Xu and R. Niu, "Displacement prediction of baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in three gorges area, china," *Computers and geosciences*, vol. 111, pp. 87–96, 2018.

[40] H. Jiang, Y. Li, C. Zhou, H. Hong, T. Glade, and K. Yin, "Landslide displacement prediction combining lstm and svr algorithms: A case study of shengjibao landslide from the three gorges reservoir area," *Applied Sciences*, vol. 10, no. 21, p. 7830, 2020.

[41] X. Zhang, C. Zhu, M. He, M. Dong, G. Zhang, and F. Zhang, "Failure mechanism and long short-term memory neural network model for landslide risk prediction," *Remote Sensing*, vol. 14, no. 1, p. 166, 2021.

[42] H. Shafizadeh-Moghadam, R. Valavi, H. Shahabi, K. Chapi, and A. Shirzadi, "Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping," *Journal of environmental management*, vol. 217, pp. 1–11, 2018.

[43] K. Khosravi, H. Shahabi, B. T. Pham, J. Adamowski, A. Shirzadi, B. Pradhan, J. Dou, H.-B. Ly, G. Gróf, H. L. Ho, H. Hong, K. Chapi, and I. Prakash, "A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods," *Journal of Hydrology*, vol. 573, pp. 311–323, 2019.

[44] O. Rahmati, M. Panahi, Z. Kalantari, E. Soltani, F. Falah, K. S. Dayal, F. Mohammadi, R. C. Deo, J. Tiefenbacher, and D. T. Bui, "Capability and robustness of novel hybridized models used for drought hazard modeling in southeast queensland, australia," *Science of The Total Environment*, vol. 718, p. 134656, 2020.

[45] H. Li, Q. Xu, Y. He, X. Fan, H. Yang, and S. Li, "Temporal detection of sharp landslide deformation with ensemble-based lstm-rnns and hurst exponent," *Geomatics, Natural Hazards and Risk*, vol. 12, no. 1, pp. 3089–3113, 2021.

[46] Y. Dai, W. Dai, W. Yu, and D. Bai, "Determination of landslide displacement warning thresholds by applying dba-lstm and numerical simulation algorithms," *Applied Sciences*, vol. 12, no. 13, p. 6690, 2022.

[47] B. Yang, K. Yin, S. Lacasse, and Z. Liu, "Time series analysis and long short-term memory neural network to predict landslide displacement," *Landslides*, vol. 16, pp. 677–694, 2019.

[48] Y. Xing, J. Yue, and C. Chen, "Interval estimation of landslide displacement prediction based on time series decomposition and long short-term memory network," *IEEE Access*, vol. 8, pp. 3187–3196, 2019.

[49] J. Li, W. Wang, and Z. Han, "A variable weight combination model for prediction on landslide displacement using ar model, lstm model, and svm model: a case study of the xinming landslide in china," *Environmental Earth Sciences*, vol. 80, no. 10, p. 386, 2021.

[50] Y. Gao, X. Chen, R. Tu, G. Chen, T. Luo, and D. Xue, "Prediction of landslide displacement based on the combined vmd-stacked lstm-tar model," *Remote Sensing*, vol. 14, no. 5, p. 1164, 2022.

[51] Z. Lin, X. Sun, and Y. Ji, "Landslide displacement prediction model using time series analysis method and modified lstm model," *Electronics*, vol. 11, no. 10, p. 1519, 2022.

[52] S. Yang, A. Jin, W. Nie, C. Liu, and Y. Li, "Research on ssa-lstm-based slope monitoring and early warning model," *Sustainability*, vol. 14, no. 16, p. 10246, 2022.

[53] E. A. Oguz, I. Depina, B. Myhre, G. Devoli, H. Rustad, and V. Thakur, "Iot-based hydrological monitoring of water-induced landslides: a case study in central norway," *Bulletin of Engineering Geology and the Environment*, vol. 81, no. 5, p. 217, 2022.

[54] A. Joshi, J. Grover, D. P. Kanungo, and R. K. Panigrahi, "Edge assisted reliable landslide early warning system," *In 2019 IEEE 16th India Council International Conference (INDICON)*, pp. 1–4, 2019.

[55] M. Gamperl, J. Singer, and K. Thuro, "Internet of things geosensor network for cost-effective landslide early warning systems," *Sensors*, vol. 21, no. 8, p. 2609, 2021.

[56] M. T. Abraham, N. Satyam, B. Pradhan, and A. M. Alamri, "Iot-based geotechnical monitoring of unstable slopes for landslide early warning in the darjeeling himalayas," *Sensors*, vol. 20, no. 9, p. 2611, 2022.

[57] C. Wang, W. Guo, K. Yang, and X. Wang, "Real-time monitoring system of landslide based on lora architecture," *Frontiers in Earth Science*, vol. 10, p. 899509, 2022.

[58] D. El Houssaini, S. Khriji, C. Viehweger, T. Keutel, and O. Kanoun, "Location-aware iot-enabled wireless sensor networks for landslide early warning," *Electronics*, vol. 11, no. 23, p. 3971, 2022.

[59] L. Piciullo, V. Capobianco, and H. Heyerdahl, "A first step towards a iot-based local early warning system for an unsaturated slope in norway," *Natural Hazards*, vol. 114, no. 3, pp. 3377–3407, 2022.

[60] Q. A. Gian, D. T. Tran, D. C. Nguyen, V. H. Nhu, and D. T. Bui, "Design and implementation of site-specific rainfall-induced landslide early warning and monitoring system: a case study at nam dan landslide (vietnam)," *Geomatics, Natural Hazards and Risk*, vol. 8, no. 2, pp. 1978–1996, 2017.

[61] T. V. Ty, P. H. Tien, L. V. Thinh, H. T. C. Hong, C. N. Thang, D. V. Duy, N. T. An, L. Q. Anh, and N. T. Liem, "Phan tich cac yeu to anh huong den on dinh bo song: truong hop nghien cuu tai doan song cha va, tinh vinh long," *Tap chi Khoa hoc Dai hoc can Tho*, vol. 58, no. 5, pp. 14–21, 2022 (Vietnamese).

[62] P. K. Diem, D. V. Den, V. Q. Minh, and N. T. H. Diep, "Danh gia tinh hinh sat lo, boi tu khu vuc ven bien tinh ca mau va bac lieu tu 1995-2010 su dung vien tham va cong nghe gis," *Tap chi Khoa hoc Dai hoc can Tho*, no. 26, pp. 35–43, 2013 (Vietnamese).

[63] H. D. Khoa, H. T. C. Hong, T. N. Thanh, N. T. T. Lieu, T. V. Ty, C. Than, C. N. Thang, H. T. Toan, H. T. Minh, D. V. Duy, and T. Q. Ninh, "Quan trac dien bien duong bo cu lao dung bang cong nghe phan tich anh vien tham," *Tap chi Vat lieu and Xay dung-Bo Xay dung*, vol. 13, no. 2, pp. 54–58, 2023 (Vietnamese).

[64] N. T. H. Diep, N. T. Loi, and N. T. Can, "Monitoring erosion and accretion situation in the coastal zone at kien giang province," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, no. 42, pp. 197–203, 2018.

[65] "Natural resources and environment newspaper. sat lo bo song, xoi lo bo bien - van de nhuc nhoi cua dong bang song cuu long," https://baotainguyenmoitruong.vn/satlobosongxoilobobienvandenhucnhoicuadongban (in Vietnamese).

[66] N. V. Liem, N. P. Dat, B. T. Dieu, V. V. Phai, P. T. Trinh, H. Q. Vinh, and T. V. Phong, "Assessment of geomorphic processes and active tectonics in con voi mountain range area (northern vietnam) using the hypsometric curve analysis method," *Vietnam Journal of Earth Sciences*, vol. 38, no. 2, pp. 202–216, 2016.

[67] O. Ghorbanzadeh, H. Shahabi, A. Crivellari, S. Homayouni, T. Blaschke, and P. Ghamisi, "Landslide detection using deep learning and object-based image analysis," *Landslides*, vol. 19, no. 4, pp. 929–939, 2022.

[68] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, pp. 1–28, 2015.

[69] V.-H. Nhu, A. Mohammadi, H. Shahabi, B. B. Ahmad, N. Al-Ansari, A. Shirzadi, M. Geertsema, V. R. Kress, S. Karimzadeh, K. V. Kamran, W. Chen, and H. Nguyen, "Landslide detection and susceptibility modeling on cameron highlands (malaysia): A comparison between random forest, logistic regression and logistic model tree algorithms," *Forests*, vol. 11, no. 8, p. 830, 2020.

[70] P. Yariyan, S. Janizadeh, T. Van Phong, H. D. Nguyen, R. Costache, H. Van Le, B. T. Pham, B. Pradhan, and J. P. Tiefenbacher, "Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping," *Water Resources Management*, vol. 34, pp. 3037–3053, 2020.

[71] K. Madhusudhanan, J. Burchert, N. Duong-Trung, S. Born, and L. Schmidt-Thieme, "U-net inspired transformer architecture for far horizon time series forecasting," *In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland*, pp. 36–52, 2022.

[72] N. Duong-Trung, D.-M. Nguyen, and D. Le-Phuoc, "Temporal saliency detection towards explainable transformer-based timeseries forecasting," *arXiv preprint arXiv:2212.07771*, 2023.

Transformative Learning Through Augmented Reality Empowered by Machine Learning for Primary School Pupils: A Real-Time Data Analysis

Abinaya M, Vadivu G
Department of Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur, Chennai

***Abstract*—Academic performance and student engagement are constant challenges in the field of modern education. When it comes to engaging students, traditional teaching methods frequently fall short, so creative solutions are needed. The Transformative potential of Augmented Reality (AR) technology as a cutting-edge teaching strategy is examined in this study. AR presents a dynamic, immersive learning environment that has the potential to completely transform conventional classrooms. By incorporating AR into the curriculum, our research transforms pedagogical paradigms, closes the engagement gap, and raises academic performance through an adaptive learning system. The study reveals the complex dynamics of AR-enhanced education through thorough analysis, powerful visualizations, and significant ANOVA results (p-value=0.03). It challenges accepted educational theories and provides insights into the complex effects on learning outcomes and student engagement. This study highlights the significance of AR in educational settings and promotes its incorporation as a transformative instrument that can establish dynamic and captivating learning environments, encourage critical thinking, creativity, and early field exploration through Artificial Intelligence (AI), and ultimately mould future leaders who can succeed.**

Keywords*—*Artificial intelligence; augmented reality; adaptive learning; machine learning; transformative learning

I. Introduction

A new era of education has begun as a result of the convergence of cutting-edge technologies and traditional teaching methods in the rapidly changing environment of modern education [1]. Augmented reality (AR), which provides educational experiences that are both immersive and interactive, is at the forefront of this revolution. These experiences are designed to captivate the minds of young learners [2]. In this digital age, when students exhibit a variety of learning styles and preferences, there has never been a greater need for personalized and adaptive learning approaches [3].The research investigates the potentially revolutionary field of augmented learning by investigating how the combination of AR and advanced machine learning algorithms has the potential to radically alter the dynamic of conventional classroom settings [4]. When it comes to education, the traditional "one-size-fits-all" approach frequently falls short of adequately addressing the specific requirements of individual students [5]. By incorporating AR technology, personalized engagement goes from being a possibility to becoming a reality [6] allowing educational experiences to be tailored to match the preferred learning pace and style of each individual student. The research is based on the conviction that AR is more than just a word [7], [8]; rather, it is the key that unlocks educational opportunities that are without parallel [9], [10].

The current particular study takes a systematic approach to analyze the effects that AR had on students both before and after it was introduced into classrooms. It seeks to highlight the transformative potential of AR by conducting research into the shifts that occur in levels of academic performance, levels of knowledge retention, and levels of engagement. The process of education is given a boost in terms of its potential to be both dynamic and interactive to AR. The purpose of the study is to shed light on the significant changes that were observed in students after they engaged with educational content that incorporated AR, thereby highlighting how important it is for education to embrace technological innovations like these. As we embark on this transformative journey, our goal is to unveil innovative strategies that will empower educators, captivate students, and pave the way for a future in which education will not simply be informative but will instead be truly immersive, engaging, and profoundly enriching.

The objective of the research is

1) To carefully evaluate how AR-enhanced learning experiences affect student's levels of engagement and retention.
2) To show, through comparison with traditional teaching methods, how much better AR-enhanced learning is at meeting the needs and styles of a diverse student body.
3) To provide verifiable data and evidence-based solutions to real-world challenges in integrating AR and ML in classrooms.
4) To give legislators, tech developers, and educators intelligent information that will shape education's use of cutting-edge technologies in the future.

II. Literature Review

Augmented reality (AR) has the capacity to transform STEM education in higher learning institutions by offering immersive and interactive learning encounters. A comprehensive literature review of 45 articles revealed that AR has been employed as an instructional tool for various STEM disciplines. However, the majority of studies have concentrated on the application of AR in laboratory-based environments, with particular emphasis on biology and chemistry. Further investigation is required to create and assess AR applications for a broader array of STEM disciplines and educational settings, focusing on aspects such as design, user involvement, and cost-effectiveness [11]. AR has the potential to revolutionize education by offering immersive and interactive experiences that promote transformative learning. Machine learning (ML) has the ability to improve AR-based learning by customizing instruction, adjusting to individual student requirements, and creating fresh educational material. AR and machine learning have the potential to revolutionize education by creating personalized and transformative learning experiences that cater to the unique needs of every student [12]. According to [13], AR has the ability to bring about significant changes and improvements in education and training. AR facilitates immersive learning experiences by enhancing the visualization of intricate concepts, promoting individualized and cooperative learning. despite the obstacles posed by expenses, the declining cost of AR devices and the demand for top-notch applications suggest a bright future for the integration of AR in education. This compels researchers and practitioners to delve into its various applications.

The author in [14] highlight the significant impact that AR can have on education, emphasizing its ability to bring about profound changes. They mention several advantages of AR in education, including heightened involvement, enhanced comprehension, increased creativity, and improved collaboration. Nevertheless, they express concerns regarding issues such as expenses, the requirement for top-notch applications, and possible diversions. Educators should thoroughly investigate AR options, carefully choose appropriate applications, and strategically design integration plans in order to effectively utilize its potential for improving student learning and engagement. According to [15], AR and virtual reality (VR) have the capacity to transform education by improving engagement, comprehension, creativity, and collaboration. Nevertheless, it is imperative to tackle obstacles such as the expenses associated with devices, the requirement for outstanding applications, and the possibility of distractions. Teachers should investigate the potential of AR and VR applications, recognizing their advantages while strategically incorporating them into the curriculum to foster captivating and interactive educational experiences for students. The author in [16], emphasises the significant impact that AR can have on Education 4.0, particularly in terms of creating immersive and interactive learning experiences. The idea they have includes AR applications such as virtual laboratories, improved field trips, collaborative environments, and customised learning experiences. Despite obstacles such as the high expenses of devices and the varying quality of applications, educators are strongly encouraged to investigate the potential uses of AR, conduct thorough research on the available technology, and carefully incorporate AR into their teaching methods to improve student involvement and learning in the era of Industry 4.0. The prevalence of AR in science education (62%), according to Fidan and Tuncel's (2018) content analysis of AR in education between 2012 and 2017. The most common type of AR was marker-based (67%), with achievement (54) and attitude (46%). In 67% of the studies, there was a positive impact on student achievement. The findings highlight the potential of AR in education, particularly in science, urging educators to investigate appropriate AR types and variables while acknowledging the need for additional research across diverse subjects. AR implemented with care can significantly improve student engagement and learning experiences.

The author in [17] analysis of AR in education (2012-2017) underscores its potential, especially in science education (62%). Marker-based AR was predominant (67%), focusing on achievement (54%) and attitude (46%). While positive impacts on student achievement were common (67%), attitudes towards AR varied. Educators are encouraged to explore AR types, consider relevant variables, and integrate high-quality applications to enhance student engagement across subjects, emphasizing the need for thoughtful planning and implementation in educational contexts [18] highlight AR potential to improve education through increased engagement, comprehension, creativity, and collaboration. Despite challenges such as device costs and application quality, educators are encouraged to investigate the benefits of augmented reality, research appropriate technologies, and integrate high-quality applications. AR can revolutionise education by creating more engaging, effective, and personalised learning experiences for all students with careful planning and implementation [19] advocate AR in medical education for immersive learning experiences, improved understanding, enhanced practical skills, and reduced costs. Despite challenges like expenses and distractions, AR offers engaging, effective, and affordable learning opportunities. Medical schools should integrate AR thoughtfully, raise faculty awareness, and conduct ongoing research to maximize its benefits for students' education.

In their paper, [20] argue that AR can transform education by improving engagement, effectiveness, and accessibility. They highlight AR's immersive, interactive learning experiences and personalised feedback and support. However, device costs, high-quality app development, and teacher training must be addressed. Teachers should investigate AR's benefits, research relevant technologies, and carefully plan integration. Safety, access, and assessment methods should be considered to maximise AR's impact on student learning. Careful planning and implementation can make AR a powerful educational tool.With the potential to provide students with immersive and interactive learning experiences that surpass the constraints of conventional teaching methods, augmented reality (AR) has the potential to revolutionise the primary school education system. The author in [21], [22] are just two of the many studies that have shown how AR improves student learning outcomes. discovered that AR-based virtual reality walkthroughs can effectively reduce motion sickness and improve learning outcomes. Furthermore, it has been demonstrated that AR-based instructional digital mo. AR technology has also been used to produce interesting experiences, such as AR-based mobile applications that assist learning in kids with learning disabilities and AR-enabled sports games that increase student motivation for physical activity [23]. AR is essential to transformative learning because it challenges students' preconceived notions and presumptions. Augmented reality (AR) provides first-hand experiences that help students develop a more nuanced worldview and a deeper understanding of history by submerging them in virtual simulations of historical events [24]. Moreover, augmented reality fosters global perspectives and tolerance in students by facilitating the exploration of various cultures and viewpoints.There are significant ramifications for elementary schools. AR has the potential to completely transform education by giving students interactive experiences that improve their comprehension of difficult concepts and by adjusting instruction to meet each student's needs [25]. AR can be used by primary school teachers to support individualised learning programmes, interactive scientific experiments, virtual field trips, and educational games. Access to AR devices, top-notch applications, and careful planning to match AR activities with the curriculum are necessary for a successful integration, though [26], [27]. AR is a potent tool that can improve learning outcomes and student engagement in elementary schools when used strategically.

A. Overcoming the Limitations of Previous Research

For a more thorough and influential study, the constraints of the previous work on AR in education are tackled in the current study. The main limitation of the previous studies are, there is a distinct focus on AR applications in science education, namely in the fields of biology and chemistry, this provide the research gap for the investigation of AR in a wider range of STEM disciplines and educational contexts. Furthermore, although the revolutionary potential of augmented reality (AR) is acknowledged in the previous study, there is a shortage of thorough investigation into the cost effectiveness, involvement of users, and design of AR applications. In order to get beyond these restrictions, the study ought to expand its scope to cover a wider variety of STEM fields and educational settings, guaranteeing a thorough assessment of AR's suitability. Along with providing useful answers to these problems, the study must also carry out a thorough investigation into the potential difficulties brought on by expenses, device costs, and application quality. In order to offer a comprehensive understanding of AR adoption in education, the study also ought to investigate the design features of AR applications, taking user involvement and cost-effectiveness into account. In order to effectively shape the future of education, educators, legislators, and tech developers access the verifiable data, evidence-based solutions, and intelligent information. By addressing these limitations, the study can provide a more delicate and practical perspective on the integration of AR in education.

TABLE I. No. of Participants Detail in the Study

GRADE	BOYS	GIRLS
GRADE 2	32	31
GRADE 3	31	33
TOTAL	63	64

III. Methods

A. Data Gathering

Grade 2 and 3 students at Little Scholars Matriculation Higher Secondary School (LSMS) in Thanjavur District, Tamil Nadu, India were selected to participate. The study included a total of 127 students, with 63 boys and 64 girls. Table I shows the grade breakdown of the participants. A variety of data points were methodically gathered and organized to guarantee thorough insight into the AR learning experience. Fig. 1 depicts the Grade 2 Children saw the Marker based AR scene in a Mobile Phone.

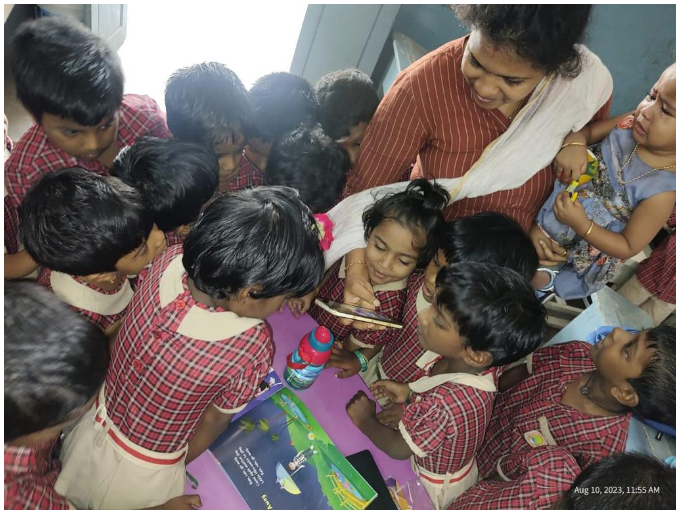


Fig. 1. Grade 2 children viewing the augmented reality scene in mobile phone.

1) Pre- and Post-Assessment: The first round of test scores was collected from the student's permanent files to reveal their knowledge levels at the outset. Before participating in the AR interventions, students underwent a pre-assessment test to evaluate their initial understanding. Students were given a starting point in the form of these pre-test scores. Subsequently, students engaged in a series of five AR-based activities designed to enhance their understanding of the subject matter. A post-task evaluation was given after participants had finished these exercises. Scores on assessments given after AR treatments were implemented were used to draw comparisons. Using this method, we were able to compare students' pre- and post-assessment scores representing their prior knowledge and the impact of AR-based learning activities, respectively to determine how much growth each student had made.

2) No. of Augmented Reality Activities: Student participants in the study used five different AR applications designed to facilitate interactive learning about fruits, animals, careers, rhymes, and the solar system. These exercises were carefully designed to stimulate deeper thought and greater participation from students. Time spent in each AR app by each student was meticulously recorded during each AR session. A built-in timer within the augmented reality applications tracked how long students spent on each activity. The information was instrumental in determining how long students were fully engaged in their studies. It also revealed important information about the level of interest that students had in the material being taught.

3) Educator Engagement and Ethics in Augmented Reality: All participant's rights and safety were protected throughout our study's entire data collection process, which was conducted in accordance with the highest ethical standards. Informed consent was diligently obtained from both the students and their guardians, signifying their voluntary agreement to participate in the AR learning activities. Our research would not have been possible without this open and ethical methodology, which emphasized the value of informed consent and personal autonomy.

An attentive teacher was there to help every step of the way, which greatly boosted the effectiveness of the AR lessons. The teacher's presence was essential to the success of the activities and contributed to a positive and encouraging classroom climate. Students benefited greatly from this direction because it made them feel safe and encouraged them to fully immerse themselves in the augmented environment.

4) Consent from Institution: Prior to the start of the study, parents received detailed consent forms outlining the nature of augmented reality activities, their benefits, and the potential educational outcomes. Only children whose parents agreed to participate in the study were included

5) Assent from Student: Students were briefed on the activities and given the option to participate if they desired. Getting people's Permission to take part in AR sessions was a precondition for doing the activity.

B. Augmented Reality Implementation

AR requires a deliberate and immersive approach in order to be integrated into the educational framework. The creation and incorporation of AR content were carefully planned to meet the goals of the curriculum and provide young students with an enjoyable learning environment.

1) AR Content Development and Integration: We took a forward-thinking approach to creating AR content with the intention of giving students access to engaging and thought-provoking educational opportunities. Our AR content's marker-based interface was both easy to use and straightforward. ThE technology made it easy for students to interact with the material, creating a stimulating classroom atmosphere. The AR modules we've developed feature vibrant 3D models, captivating animations, and enveloping auditory cues, all of which were thoughtfully developed. Students' interest and understanding of the material were both boosted by the incorporation of these interactive features. Modules about fruits, vegetables, careers, the solar system, and rhymes were developed. Each AR module was created with a distinct set of pedagogical goals in mind. For instance, fruits and vegetables taught students not only about agriculture and nutrition but also about the value of caring for one's body and the environment. the Professions AR content encouraged early exploration and understanding of a variety of professions by introducing children to them. Students' minds were blown by the solar system modules, which took them on a fascinating journey through space. The rhymes unit also made use of visual aids, which helped to develop both vocabulary and imagination.

2) AR Software and Platforms: The merging of Unity, an advanced game development engine, and Vuforia, an innovative AR platform, has resulted in a revolutionary era of interactive learning. The ground-breaking collaboration gave rise to Marker-based AR, which completely altered the educational landscape. Unlike conventional markers or triggers, this technology can instantly identify physical books and printed materials, triggering digital overlays within the AR app with ease. The app's intuitive design makes it suitable for elementary school students, and its AR content is accessible to even the youngest students. Integrating five separate educational activities studying fruits and vegetables, animals, careers, nursery rhymes, and the solar system improves participation and memorization. Through the use of school issued tablets and our own mobile devices, students are able to access these enlightening modules outside of the confines of the classroom.

C. Algorithm

1. Collect the pre assessment score of students
2. Analysis of Augmented Reality Interactions with post assessment score, time of interactions and the no of activities completed and the improvement score
3. Preprocess the collected Data
4. Selection of Features
5. Machine Learning in Statistical Analysis
6. Analyse the statistical analysis results
7. Use graphs and charts to visually represent findings

IV. Results

AR technology has brought in a new era of immersive learning at LSMS School, as seen in the Fig. 1 interesting photo of students participating in an AR-enhanced lesson. Their reactions, full of wonder and teamwork, reflect the fresh energy that AR has brought to their academic pursuits. Fig. 2, a Box Plot showing before and after AR scores, vividly illustrates how this excitement translated into remarkable academic progress. Students' impressive progress after adopting AR-based learning experiences is effectively demonstrated by this visual representation.

A visualization that captures how AR has transformed student performance. The correlation between AR enhanced learning experiences and subsequent improvements in students' scores is clearly displayed by the graph. Our box plot, which displays a compelling comparison of pre and post-augmented reality scores, is a testament to the thorough analysis carried out. The distinct division between the two stages highlights the noteworthy advancements that students achieved following the incorporation of augmented reality-based educational opportunities. The impactful, precise data representation reveals the various ways in which students reacted to AR interventions by highlighting not only the mean scores but also their distribution and spread. The Violin Plot in the Fig. 3 provides a more comprehensive look at how AR activities correlate with assessment results. Its graceful arcs and peaks represent the wide range of reactions from students at different engagement levels. Like individual notes in an educational piece, each activity tally represents a distinct educational opportunity. Increases in the violin's dynamic range signify collective growth wherein agreed-upon tasks significantly advance shared goals. Numerous differences denote unique learning paths, necessitating a pliable instructional strategy analogous to a skilled conductor leading a meticulously prepared band. This complex web of relationships exemplifies the individualised nature of education that caters to different approaches to learning.

Incorporating AR into educational practices is effective, as evidenced by the observed increase in improvement scores and

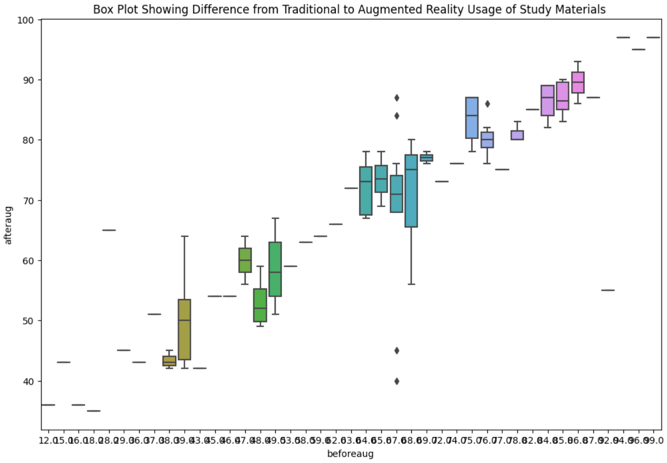


Fig. 2. Box plot shows the pre and post marks of the students after using augmented Reality in to their curriculum activity.

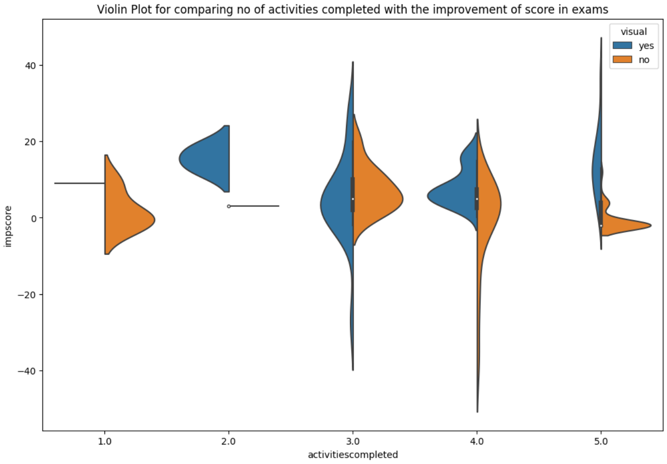


Fig. 3. Violin plot shows the no.of activities in augmented reality increases the marks in exam.

the significant p-value of 0.03 from ANOVA analysis when compared with the pre- and post-assessment scores of the students. The result confirms that AR not only improves understanding and interest, but also facilitates unique educational paths. The findings point to a future where technology and education work hand in hand to provide each student with a stimulating and individualised education. the integration of AR tools with conventional classroom practices has revolutionised education. Students, once confined by conventional boundaries, now explore a spectrum of knowledge, fostering a harmonious blend of technology and education. These results not only prove that AR interventions work, but also stress the importance of using flexible methods of instruction to maximise the benefits of this new paradigm in education.

The scatter plot in the Fig. 4 results clearly shows the complex relationship between students' improvement scores and their augmented reality experiences. Plotting the individual student data points reveals an interesting pattern that highlights the beneficial effects of augmented reality on academic advancement. The values that are systematically extracted from

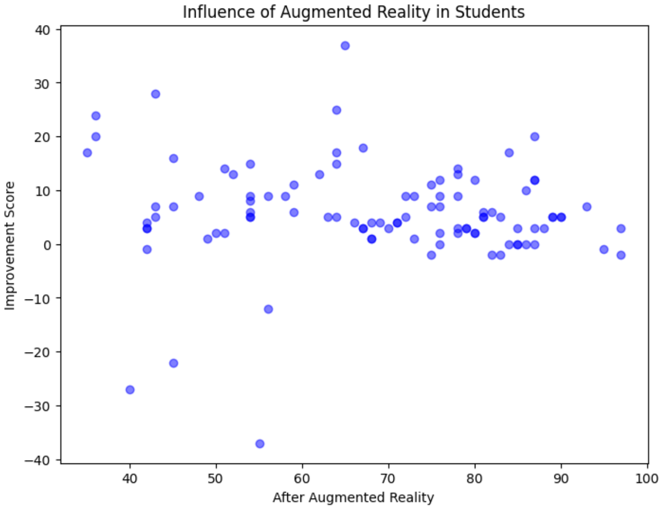


Fig. 4. Scatter plot to visualize the improvement of marks in students after using augmented reality.

the augmented reality interventions and improvement scores provide an inspiring tale that highlights the revolutionary potential of cutting-edge educational technologies.

V. Discussion

Fig. 1 effectively captures the immersive quality of augmented reality, where traditional barriers to education vanish and interactive, hands-on learning becomes possible. Pupils can be seen examining virtual objects and even reaching out to touch and interact with them. Their obvious excitement reflects a deeper level of engagement made possible by augmented reality technology. This dynamic interaction turns abstract ideas into real, memorable experiences that go beyond the pages of textbooks.Beyond the technical marvel, the picture emphasises how AR fosters a collaborative spirit. In the augmented environment, students are seen collaborating to solve problems, sharing insights, and highlighting discoveries. In addition to improving their comprehension, this cooperative learning environment helps them develop critical thinking, communication, and teamwork skills.This graphic insight highlights the critical role that augmented reality plays in boosting student engagement and encouraging collaborative learning experiences. The excitement that the students displayed is consistent with the favourable results that our study showed, confirming AR's potential as a revolutionary teaching tool.

the determination, curiosity, and participation of the pupils. It represents an evolution in the learning process rather than just an improvement in scores. With AR, education becomes more than just a teaching tool; it becomes a shared experience in which teachers and students work together in a dynamic, immersive learning environment. Upon considering the effects of this Fig. 5, it is evident that AR is more than a superficial trend; rather, it is a revolutionary force influencing the course of education. It has the infinite capacity to captivate, motivate, and enhance educational experiences. In a larger sense, this visual witness questions established educational theories and calls on establishments to adopt cutting-edge technologies and reconsider the definition of a classroom.

The plot also presents the idea of outliers, or students whose performance increased after AR was implemented. These outliers are more than just statistics; in which augmented reality served as a catalyst to enable remarkable academic achievements. These anomalies cast doubt on accepted ideas about learning paths, highlighting the revolutionary potential of immersive learning technologies. There are significant consequences to be taken from this box plot. It highlights that AR is acustomized experience that accommodates a range of learning styles and speeds rather than an inflexible tool. This necessitates a change in our pedagogical paradigms as educators—from standardized instruction to individualized, AR-enhanced learning. It calls for the development of flexible, responsive curricula that value each student's uniqueness and promote an atmosphere in which unusual growth is not only welcomed but encouraged.

Unexpectedly, the plot also presents the idea of outliers, or students whose performance increased after AR was implemented. These outliers are more than just statistics; in which AR served as a catalyst to enable remarkable academic achievements. These anomalies cast doubt on accepted ideas about learning paths, highlighting the revolutionary potential of immersive learning technologies. The ANOVA test yielded a significant p-value of 0.03 that is highly significant. According to statistics, there is less than a 5 per cent chance that the observed differences in improvement scores between traditional and augmented reality methods are the result of random fluctuations if the p-value is less than the conventional threshold of 0.05. Because of this low p-value, the idea that augmented reality is a key component enhancing learning outcomes is strongly supported.

A strong degree of confidence in the observed results is indicated by the 0.03 p-value, which supports the finding that augmented reality is significantly linked to higher student test scores. This statistical result emphasises the validity and reliability of the study's findings and is consistent with our qualitative observations and analyses.

This confirmation strengthens the case for the use of AR technologies in teaching methods, as does the accompanying picture that shows student participation and visual proof.

The compelling picture, showcasing students' increased enthusiasm and active participation in AR classes, is a perfect visual representation of our study's striking findings. This deeper involvement with the material explains why students are performing better in school, underscoring the significant effect AR has on education. AR's interactive nature enables immersive and individualised learning, meeting the needs of students with a wide range of learning styles. We believe that students' increased flexibility is a major factor in the positive impact on their academic growth that we have observed. The visually impressive documentation supports the centrality of AR in education, which is emphasised by our analysis. The classroom environment is improved and students learn more when technology is used in the classroom. We believe that AR can be a game-changer in the classroom by creating more interactive and collaborative learning spaces. Our AR content has been highly commended for its ability to foster critical thinking, creativity, and early exploration across a wide range of disciplines. This cutting-edge method guarantees that students not only learn new information but also develop crucial skills necessary for their future success.

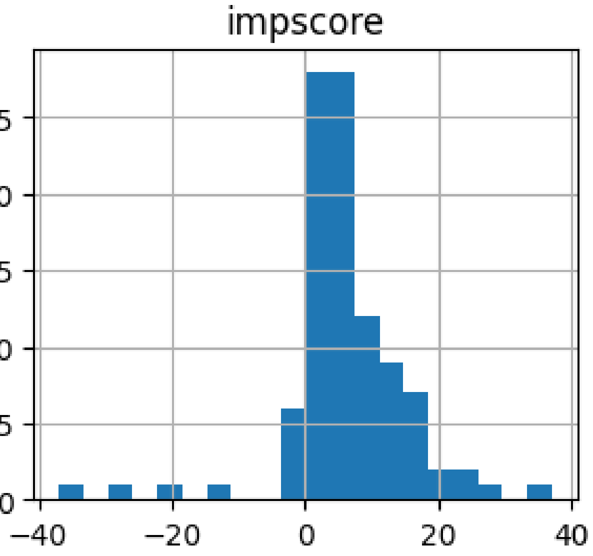


Fig. 5. Statstical representation of overall improvement score of the students.

VI. Conclusion

As an outcome, the statistical evaluation highlights the effect of AR on the academic achievement of learners, as evidenced by a significant ANOVA result with a p-value of 0.03. The result validated the usefulness of AR in educational settings by showing that the variation in improvement scores between traditional and AR methods .AR is crucial to improve learning outcomes was supported by the sample data collected from the school, which includes powerful diagrams comparing pre- and post-augmented reality scores. The box plot, which clearly contrasts the use of AR with traditional methods, offers a clear visual depiction of the improvement in student performance. The positive correlation between interactive learning experiences and academic progress is further supported by the violin plot, which correlates the number of activities with improvement scores. our study explored the practical implications of augmented reality in addition to establishing its statistical significance. The image, which shows students' enthusiastic participation and active engagement during augmented reality sessions, is a potent illustration of how this technology has revolutionised the classroom and the results demonstrate that augmented reality is an initiator for a pedagogical revolution and not simply a technical development. Adopting cutting-edge technologies like AR is essential as education changes because it allows teachers to design dynamic, personalized, and interactive learning environments. It is recommended that educational institutions and instructors incorporate augmented reality into their pedagogical approaches, capitalizing on its capacity to enhance the learning experience and equip learners for a technologically and knowledge-driven future.

VII. Future Gap

Our research sheds light on important developments in the field of AR-enhanced education, but there are still many questions that need to be answered. Long-term research into effects is an exciting avenue to explore. The longer the study runs, the more insight we'll have into the long-term effects of AR on students' learning trajectories and retention rates. Exploring AR's utility and efficacy in a variety of subject areas and with students of varying ages could shed light on the technology's potential in the classroom as a whole. Furthermore, investigating how best to combine AR with other pedagogical approaches like gamification or individualized learning could lead to novel and complementary learning strategies. Ethnographic studies could provide detailed insights into students' experiences, shedding light on the emotional and social dimensions of augmented reality-enhanced learning. In the final analysis foremost, addressing the current gaps in technology access by investing in the development of low-cost and easily accessible augmented reality tools for resource-constrained educational settings would guarantee an inclusive educational landscape. By filling these future gaps, we can improve AR's use in the classroom and create more engaging, effective, and equitable learning environments for all students.

ACKNOWLEDGMENT

The Full-time Ph.D. - IT/IT Enabled Services (IT/ITES) is supported by the Ministry of Electronics and Information Technology, Government of India initiated "Visvesvaraya Ph.D. Scheme for Electronics and IT". Funding Statement: The authors received no specific funding for this study and also thanks to Little Scholars Matriculation Higher Secondary School, Thanjavur District, Tamil Nadu, India.

REFERENCES

[1] Hanson, A. H., Krause, L. K., Simmons, R. N., Ellis, J. I., Gamble, R. G., Jensen, J. D., ... & Dellavalle, R. P. (2011). Dermatology education and the Internet: traditional and cutting-edge resources. *Journal of the American Academy of Dermatology*, *65*(4), 836-842.

[2] Avila-Garzon, C., Bacca-Acosta, J., Duarte, J., & Betancourt, J. (2021). Augmented Reality in Education: An Overview of Twenty-Five Years of Research. *Contemporary Educational Technology*, *13*(3).

[3] Khosravi, H., Sadiq, S., & Gasevic, D. (2020, February). Development and adoption of an adaptive learning system: Reflections and lessons learned. In *Proceedings of the 51st ACM technical symposium on computer science education* (pp. 58-64).

[4] Asif, M. A., Al Wadhahi, F., Rehman, M. H., Kalban, I. A., & Achuthan, G. (2020). Intelligent educational system for autistic children using augmented reality and machine learning. In *Innovative Data Communication Technologies and Application: ICIDCA 2019* (pp. 524-534). Springer International Publishing.

[5] Taft, S. H., Kesten, K., & El-Banna, M. M. (2019). One Size Does Not Fit All: Toward an Evidence-Based Framework for Determining Online Course Enrollment Sizes in Higher Education. *Online Learning*, *23*(3), 188-233.

[6] Lee, Y., & Lee, C. H. (2018). Augmented reality for personalized nanomedicines. *Biotechnology advances*, *36*(1), 335-343.

[7] Hu, X., Goh, Y. M., & Lin, A. (2021). Educational impact of an Augmented Reality (AR) application for teaching structural systems to non-engineering students. *Advanced Engineering Informatics*, *50*, 101436.

[8] Yilmaz, O. (2021). Augmented Reality in Science Education: An Application in Higher Education. *Shanlax International Journal of Education*, *9*(3), 136-148.

[9] Kim, Y., Chaivisit, S., Tutaleni, A., & Stansberry, S. (2019, March). An Exploratory Research on Emerging Technologies in a Transformative Learning Space. In *Society for Information Technology & Teacher Education International Conference* (pp. 1880-1887). Association for the Advancement of Computing in Education (AACE).

[10] Riva, G., Baños, R. M., Botella, C., Mantovani, F., & Gaggioli, A. (2016). Transforming experience: the potential of augmented reality and virtual reality for enhancing personal and clinical change. *Frontiers in psychiatry*, *7*, 164.

[11] Mystakidis, S., Christopoulos, A., & Pellas, N. (2022). A systematic mapping review of augmented reality applications to support STEM learning in higher education. *Education and Information Technologies*, *27*(2), 1883-1927.

[12] Cabiria, J. (2012). Augmenting engagement: Augmented reality in education. In *Increasing student engagement and retention using immersive interfaces: Virtual worlds, gaming, and simulation* (pp. 225-251). Emerald Group Publishing Limited.

[13] Lee, K. (2012). Augmented reality in education and training. *TechTrends*, *56*, 13-21.

[14] Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented Reality in education–cases, places and potentials. *Educational Media International*, *51*(1), 1-15.

[15] Elmqaddem, N. (2019). Augmented reality and virtual reality in education. Myth or reality?. *International journal of emerging technologies in learning*, *14*(3).

[16] Martin, J., Bohuslava, J., & Igor, H. (2018, September). Augmented reality in education 4.0. In *2018 ieee 13th international scientific and technical conference on computer sciences and information technologies (CSIT)* (Vol. 1, pp. 231-236). IEEE.

[17] Fidan, M., & Tuncel, M. (2018). Augmented Reality in Education Researches (2012-2017): A Content Analysis. *Cypriot Journal of Educational Sciences*, *13*(4), 577-589.

[18] Kraut, B., & Jeknić, J. (2015, May). Improving education experience with augmented reality (AR). In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 755-760). IEEE.

[19] Kamphuis, C., Barsom, E., Schijven, M., & Christoph, N. (2014). Augmented reality in medical education?. *Perspectives on medical education*, *3*, 300-311.

[20] De Lima, C. B., Walton, S., & Owen, T. (2022). A critical outlook at augmented reality and its adoption in education. *Computers and Education Open*, 100103.

[21] M. D. A. Anua, I. Ismail, N. S. M. Shapri, W. M. A. F. W. Hamzah, M. M. Amin, and F. Karim, "Validate the Users' Comfortable Level in the Virtual Reality Walkthrough Environment for Minimizing Motion Sickness," ProQuest, vol. Volume 14, no. Issue 4, 2023.

[22] N. Yahya et al., "Instructional Digital Model to Promote Virtual Teaching and Learning for Autism Care Centres," IJACSA) International Journal of Advanced Computer Science and Applications, vol. 14, no. 6, 2023, Accessed: Jul. 03, 2023.

[23] B. Doskarayev et al., "Development of Computer Vision-enabled Augmented Reality Games to Increase Motivation for Sports," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 14, no. 4, 2023.

[24] M. Lazo-Amado, L. Cueva-Ruiz, and L. Andrade-Arenas, "Designing a Mobile Application using Augmented Reality: The Case of Children with Learning Disabilities," International Journal of Advanced Computer Science and Applications, vol. 13, no. 6, 2022.

[25] J. Qu, B. Ma, L. Zheng, and Y. Kang, "Design and Implementation of Teaching Assistant System for Mechanical Course based on Mobile AR Technology," International Journal of Advanced Computer Science and Applications, vol. 13, no. 4, 2022.

[26] Z. R. Mahayuddin and A. F. M. S. Saif, "Vision based 3D Gesture Tracking using Augmented Reality and Virtual Reality for Improved Learning Applications," International Journal of Advanced Computer Science and Applications, vol. 12, no. 12, 2021.

[27] S. A. H. Morales, L. Andrade-Arenas, A. Delgado, and E. L. Huamani, "Augmented Reality: Prototype for the Teaching-Learning Process in Peru," International Journal of Advanced Computer Science and Applications, vol. 13, no. 1, 2022.

Indonesian Twitter Emotion Recognition Model using Feature Engineering

Rhio Sutoyo[1], Harco Leslie Hendric Spits Warnars[2], Sani Muhamad Isa[3], Widodo Budiharto[4]
Computer Science Department-BINUS Graduate Program-Doctor of Computer Science,
Bina Nusantara University, Jakarta, Indonesia 11480[1,2]
Computer Science Department-BINUS Graduate Program-Master of Computer Science,
Bina Nusantara University, Jakarta, Indonesia 11480[3]
Computer Science Department-School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480[4]

Abstract—**Twitter is a social media platform that has a large amount of unstructured natural language text. The content of Twitter can be utilized to capture human behavior via emphasized emotions located in tweets. In their tweets, people commonly express emotions to show their feelings. Hence, it is crucial to recognize the text's underlined emotions to understand the message's meaning. Feature engineering is the process of improving raw data into often overlooked features. This research explores feature engineering techniques to find the best features for building an emotion recognition model on the Indonesian Twitter dataset. Two different text data representations were used, namely, TF-IDF and word embedding. This research proposed 12 feature engineering configurations in TF-IDF by combining data stemming, data augmentation, and machine learning classifiers. Moreover, this research proposed 27 feature engineering configurations in word embedding by combining three-word embedding models, three pooling techniques, and three machine-learning classifiers. In total, there are 39 feature engineering combinations. The configuration with the best F_1 score is TF-IDF with logistic regression, stemmed dataset, and augmented dataset. The model achieved 65.27% accuracy and 66.09% F_1 score. The detailed characteristics from the top seven models in TF-IDF also follow the same feature engineering configuration. Lastly, this work improves performance from the previous research by 1.44% and 2.01% on the word2vec and fastText approaches, respectively.**

Keywords*—*Text classification; feature engineering; emotion recognition; Indonesian tweet; natural language processing

I. Introduction

Twitter is a social media platform that provides services to share ideas and opinions. The popularity of Twitter leads to millions of users sending data in the form of tweets, i.e. a short message from Twitter users [1]. As a result, Twitter has a large amount of unstructured natural language text. Researchers have used the content of Twitter to predict economic trends, e.g., financial market prediction [2]. Furthermore, Twitter data can also be utilized to capture human behavior via emphasized emotions located in tweets.

Based on Shaver's theory, emotions can be categorized into five basic classes, i.e., anger, fear, happiness, love, and sadness [3]. On Twitter, people commonly express emotions to show their feelings toward something, e.g., political party, sexual abuse, or simply a bitter experience in their life. Thus, it is crucial to recognize the text's underlined emotion to understand the message's meaning [4]. The underlined emotions can be identified directly via emotion words, e.g., *dukacita* (grief) for sadness, and *kesal* (annoyed) for anger. However, Twitter users can also implicitly display emotions in their tweets, making them hard to identify.

The ability to recognize emotions automatically is essential for various applications. In politics, emotion recognition can predict the polarity of the sentiment in the Presidential election based on social media data [5]. Emotion recognition can also be utilized for games with emotion-based dynamic difficulty adjustment [6]. Furthermore, emotion recognition can also be utilized to build a recommendation system for culinary and food [7]. Lastly, the emotion model can be used to build an emotionally aware chatbot capable of recognizing and interpreting human emotions [8].

There are several challenges to building an automatic emotion recognition model. First, the natural language text in Twitter data is unstructured and uncontrolled, e.g., the length of tweets might be too short or too long, the texts contain typos, and the texts contain misused terms. Second, Twitter datasets for emotion recognition tasks are primarily available for the English language [9], [10]. Datasets for Indonesian emotion recognition from previous studies are not available publicly [11], [12]. Fortunately, Saputri et al. share their Indonesian Twitter dataset for emotion classification task [13].

Inspired by the work of [13], the authors explore feature engineering to build an emotion recognition model on the Indonesian Twitter dataset. This work extends the work of [13] by further exploring feature engineering techniques to find the best features for identifying emotion in Indonesian Tweet. It combines various data preprocessing techniques, word embedding models, different pooling techniques, and machine learning classifiers. The limitation of this work is it does not include an experiment using a combination of features. Specifically, this work focuses on basic features, namely, Bag-of-Words (BOW), Word2Vec (WV), and FastText (FT).

In total, 39 feature engineering configurations are proposed and compared. This work compares the best feature engineering configurations based on the F_1 score metric for evaluation. The best F_1 score achieved was 66.09% from the TF-IDF approach with logistic regression, stemmed dataset, and augmented dataset. The experiment results have shown that the TF-IDF text representation is better than word embedding. On average, the stemming process increases the accuracy performance by 0.11%. Nevertheless, it decreases the F_1 score by 0.06%. Moreover, the dataset augmentation reduces both accuracy and F_1 score by 0.45% and 0.67%, respectively. These results contribute to comprehensively exploring feature engineering techniques to identify the best features for building emotion recognition models on the Indonesian Twitter dataset [13]. It also has potential practical implications for developing an emotionally aware chatbot capable of recognizing human emotions.

The structure of this paper is as follows: related works are presented in Section II. The system architecture of this work is described in Section III. The results and discussions are presented in Section IV and Section V, respectively. Lastly, the conclusion and future work of this work are discussed in Section VI.

II. Related Works

A. Emotion Recognition

Emotion recognition is an attempt to recognize and classify people's emotions. The basic pipeline of emotion recognition is data preprocessing, data representation, training model, and model evaluation. There are two superior-level categories of emotions: positive and negative [3]. Two major basic level categories in the positive category were love and happiness. The three major basic level categories in the negative category were anger, fear, and sadness. Emotion recognition tasks can be performed in various modalities, such as textual [13], audio [14], speech [15], and brain activity [16]. Furthermore, it can also be performed in a multimodal dataset [4].

B. Dataset of Emotion Recognition

Several datasets can be utilized to train the emotional recognition model. The ISEAR [17], the Tales [18], and the AffectiveText [19] are known datasets that are available in the English language. The ISEAR consists of 7,665 sentences labeled with a specific emotion, i.e., joy, fear, anger, sadness, disgust, shame, and guilt [17]. The Tales consists of 15,302 sentences from 176 stories by three different authors [18]. It utilizes Ekman's six basic emotions theory [20], merging anger and disgust. The AffectiveText consists of 1,250 instances from news headlines and has six basic emotions models of Ekman complemented by its valence [19]. Furthermore, the emotions recognition dataset is also available in the Indonesian language, i.e. the Indonesian Twitter Emotion Dataset [13]. It consists of 4,403 Indonesian tweets with general topics. The emotions model is labeled using Shaver's five basic emotions: love, anger, sadness, and fear [3]. The summary of the emotion recognition dataset is listed in Table I.

The study of [13] has performed feature engineering using various text features: lexicon-based, Bag-of-Words, word embeddings, orthography, and Part-of-Speech (POS) tags. In the experiment, [13] has represented word embeddings with several dimensional variations. However, the research has not focused on exploring feature engineering with various word embedding pooling techniques. This paper addresses that issue by exploring various feature engineering configurations, including using various word embedding pooling techniques, to build an Indonesian emotion recognition model.

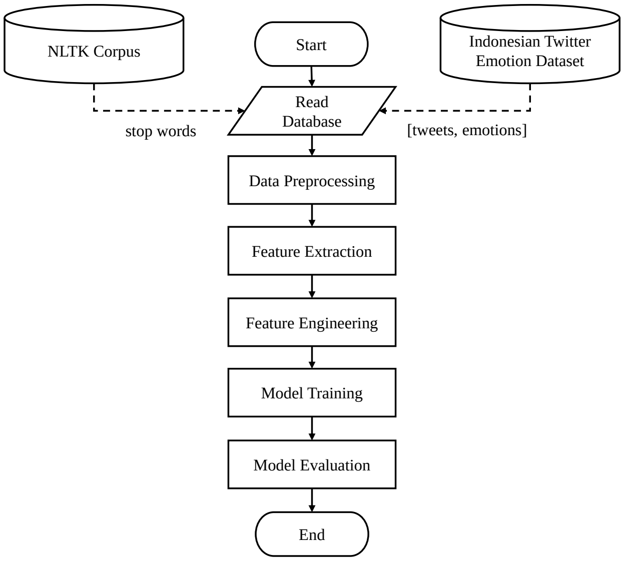


Fig. 1. The system architecture of Indonesian twitter emotions recognition using feature engineering.

III. System Architecture

This work uses feature engineering to present a system architecture for Indonesian Twitter emotion recognition (see Fig. 1). There are two sources of the dataset: NLTK Corpus[1] for a list of Indonesian stop words and Indonesian Twitter Emotions Dataset [13] for model training.

The arrow in the system architecture shows the direction of the process. Furthermore, the dashed arrow shows the data flow. The details of each step in building our model are explained in the following sections:

A. NLTK Corpus

In data preprocessing, the system removes stop words from text inputs. Stops words are low-value words that generally do not contain helpful sentence meanings. The library of NLTK Corpus contains approximately 750 Indonesian stop words. Examples of stop words are *adalah* (is), *agak* (somewhat), and *ke* (to).

B. Indonesian Twitter Emotion Dataset

This work utilizes the Indonesian Twitter emotion dataset from [13] for model training. Saputri et al. collected 4,403 Indonesian tweets using Twitter Streaming API for two weeks. The Twitter metadata, namely, username, hyperlink, and phone number in the sentences, are converted into special tags (i.e., [USERNAME], [URL], and [SENSITIVE-NO]). Then, they annotated the collected dataset with Shaver's emotions model [3]. The emotion distribution of the dataset is shown in Fig. 2.

[1] https://www.nltk.org/api/nltk.corpus.html

TABLE I. PUBLICLY AVAILABLE EMOTION RECOGNITION DATASET

No	Dataset Name	Size	Granularity	Language
1	ISEAR [17]	7,665	descriptions	English
2	Tales [18]	15,302	sentences	English
3	AffectiveText [19]	1,250	headlines	English
4	Indonesian Twitter Emotion Dataset [13]	4,403	tweets	Indonesian

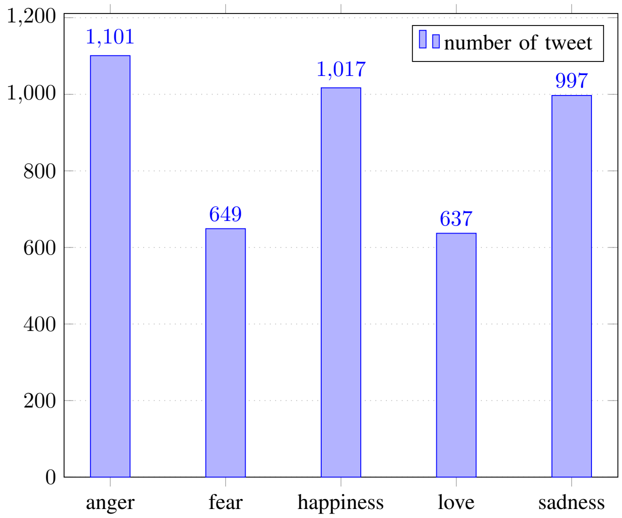


Fig. 2. The class distribution of Indonesian twitter emotion dataset.

C. Read Database

This section explains the process of reading and utilizing data from the data sources. The NLTK corpus was utilized to get Indonesian stop words. Then, the list of stop words was used in the data preprocessing step. The Indonesian Twitter emotion dataset is stored in Pandas DataFrame. Then, the data were preprocessed and converted to features for the model training.

D. Data Preprocessing

Data is the training material for the emotions recognition model. The quality of the model depends on the readiness of the data. Hence, the data preprocessing prepares data for the model training by performing several improvement steps. The case-folding converts all sentences to lowercase to avoid counting the same words as different information. Irrelevant information (i.e. username, URL, sensitive number) is removed because it is unrelated to the emotion recognition task. Then, the inputs were standardized using regular expressions to keep only the alphabet a – z. The stemming algorithm converts words to their roots. For this task, the Sastrawi python library was applied, utilizing the algorithm of Nazief and Adriani [21]. The stemming process was executed in approximately 25 minutes on 17,903 tokens. Lastly, stop words were removed to focus on important words. As a result, the data preprocessing reduces the tokens to 13,521. The algorithm illustration for the data preprocessing is shown in Algorithm 1.

Algorithm 1 Algorithm for data preprocessing

```
$corpus = [\ ]$
$i \leftarrow 0$
$N \leftarrow len(tweets)$
while $i < N$ do
    $tweet \leftarrow tweets[i]$                                   ▷ take a single tweet
    $tweet = tweet.lower()$                                        ▷ case folding
    $tweet = tweet.removeIrrelevantInformation()$
    $tweet = re.sub('[\hat{}a\text{-}z]+', ' ', tweet)$            ▷ standardization
    $tweet = stemmer.stem(tweet)$                                  ▷ stemming
    $tweet = tweet.removeStopWords()$
    $corpus.append(tweet)$                                         ▷ combine tweet
end while
```

E. Feature Extraction

This work utilizes two types of word representation for text analysis: TF-IDF and word embedding. The TF-IDF represents a document as a vector. On the other hand, word embedding represents a word as a vector. Furthermore, the length of array word embedding has a fixed array length, i.e., 300 dimensions. In the TF-IDF, the array's length depends on the size of the bag-of-words (BoW).

Three models of word embedding are utilized in this work. First, word2vec consists of 129,390 vocabulary sizes with 400 sizes of vector [13]. Second, fastText consists of 69,465 sizes of vocabulary with 100 sizes of vector [13]. Lastly, the neural network language modeling (NNLM) architecture from Google has 128 vector sizes[2]. It is trained on the Indonesian Google News 3B corpus.

In the word embedding approach, each word vector from a tweet is pooled into one vector. This work utilizes three techniques: mean pooling, sum pooling, and min-max pooling. The sum pooling sums up all vectors into the pooled vector. The mean pooling sums up all vectors and then finds their average value as the pooled vector. Lastly, the min-max pooling combines the minimum vector with the maximum vector. The result of min-max pooling is double the size of a vector. The word embedding vectorization and the pooling technique are presented in Fig. 3.

F. Feature Engineering

This research performs and combines several text-processing techniques to extract features from texts for model training. Two different text data representations were used, namely, TF-IDF and word embedding. The dataset was split into a train set (80%) and a test set (20%). This work uses a stratified train-test method to ensure both sets have a proportioned class distribution. Furthermore, a random seed '88' was used to ensure reproducible results.

SMOTE was used to perform data augmentation, i.e., over-sampling the data for model training [22]. The synthetic data were created at the vector level. The data was sampled by using a maximum sampling strategy. Moreover, a fixed random seed, i.e., '88', was used consistently to get the same sampling

[2] https://tfhub.dev/google/tf2-preview/nnlm-id-dim128/1

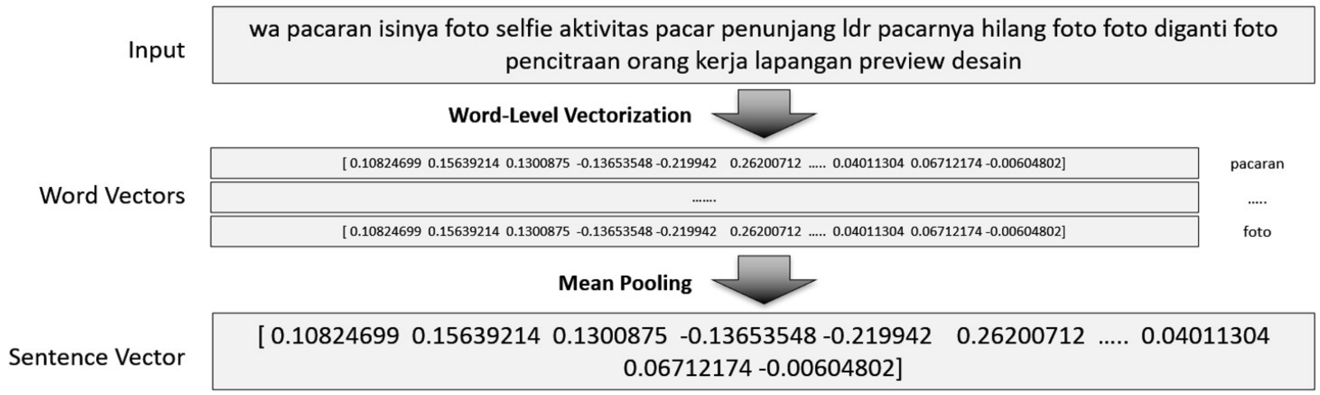


Fig. 3. The illustration of word embedding and pooling process.

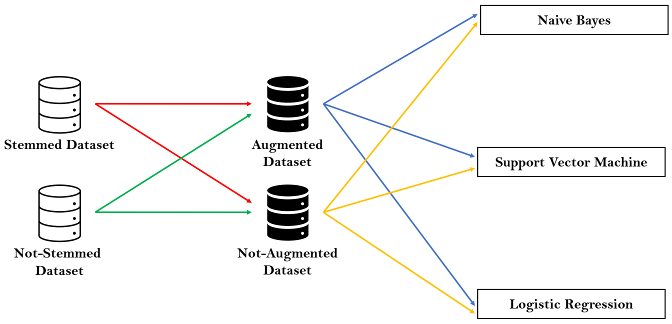


Fig. 4. Feature engineering configurations in TF-IDF.

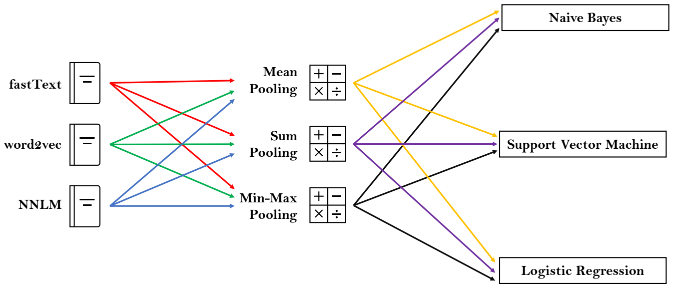


Fig. 5. Feature engineering configurations in word embedding.

result. Different machine learning classifiers were utilized for training the emotion recognition model. Based on the preliminary research, the recommended classifiers are Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM).

This research proposes twelve feature engineering configurations in TF-IDF by mixing data stemming, data augmentation, and machine learning classifiers. Fig. 4 shows the feature engineering configuration illustration for TF-IDF. The red line symbolizes the data flow of the stemmed dataset, and the green line symbolizes the data flow of the not-stemmed dataset. Then, those data were passed to the data augmentation process. The blue lines symbolize the data flow of the augmented dataset, and the yellow line symbolizes the data flow of the not-augmented dataset. Finally, those data were passed to three different classifiers.

This research proposes 27 feature engineering configurations in word embedding by mixing three-word embedding models, three pooling techniques, and three machine-learning classifiers. Fig. 5 shows the feature engineering configuration illustration for word embedding. The red, green, and blue lines symbolize feature extraction using fastText, word2vec, and NNLM, respectively. Then, the word embedding results were combined into sentence embedding with three different pooling techniques. The yellow, purple, and black lines symbolize the pooling technique using mean pooling, sum pooling, and min-max pooling, respectively. Finally, the data were passed to three different classifiers.

This work creates a naming scheme to differentiate the configurations. For example, the configuration of TF-IDF with SVM, not-stemmed database, and augmented is “TFIDF_SVM_notstem_aug”. Moreover, the configuration of word embedding with Naive Bayes, fastText, and sum pooling is “WE_NB_ft_sum”. The detail of the naming scheme is as follows:

- **TF-IDF**: TFIDF_name of classifier technique_status of dataset stemming_status of dataset augmentation.
- **Word Embedding**: WE_classifier technique_word embedding model_pooling technique.

G. Model Training

The emotion recognition model was trained using Google Colab[3]. The model training uses ’88’ as the random seed. Furthermore, the model training was performed using Python v3.7.13, NLTK v3.7, scikit-learn v1.0.2, and various text processing libraries (e.g., *re* for regular expression operations).

H. Model Evaluation

The emotion recognition model is evaluated using accuracy and macro F_1 score. The accuracy score is calculated by dividing the number of correct predictions by the number of total data. The equation of accuracy metric is presented in Equation (1).

[3] http://colab.research.google.com/

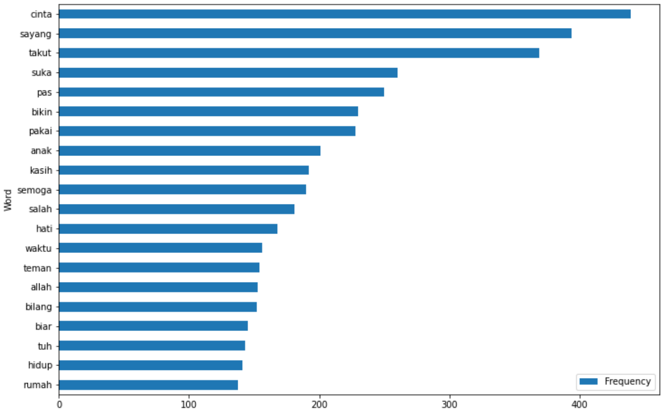


Fig. 6. Word frequency analysis of top 20 common words (excluding stop words).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

In machine learning, TP means true positive, TN means true negative, FP means false positive, and FN means false negative.

$$F_1 = \frac{2 * TP}{2 * TP + FP + FN} \tag{2}$$

The F_1 metric combines precision and recall to calculate the harmonic mean of the two metrics. The equation of F_1 metric is presented in Equation (2).

IV. Results

This section discusses the result of exploring the Indonesian Twitter emotion dataset. Furthermore, the result of the experiment is divided into two subsections, namely TF-IDF and Word Embedding. The full results of the experiment can be viewed on the GitHub page[4].

A. Dataset Exploration

Fig. 6 shows the word frequency analysis of the top 20 common words in the Indonesian Twitter emotion dataset. The result has excluded stop words that are valueless for model training. It shows the words "*cinta*" (love), "*sayang*" (darling), "*takut*" (afraid), and "*suka*" (like) are the most commonly occurring word frequencies.

Moreover, the frequent use of words is also shown as a word cloud. The diagram below includes all words from the dataset. It can be seen in Fig. 7 that the most common words are dominated by stop words, e.g., "*saya* (I)," "*yang* (which)," and "*kamu* (you)."

B. TF-IDF

This work explored 12 feature engineering configurations of TF-IDF by combining data stemming, data augmentation, and machine learning classifiers. The F_1 score results for every configuration with TF-IDF are presented in Fig. 8.

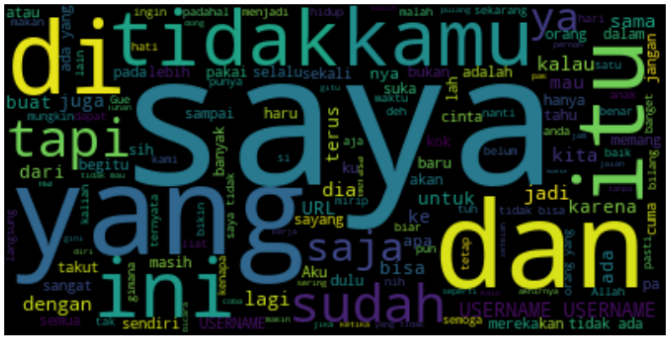


Fig. 7. Word cloud plot of Indonesian twitter emotion dataset (including all words).

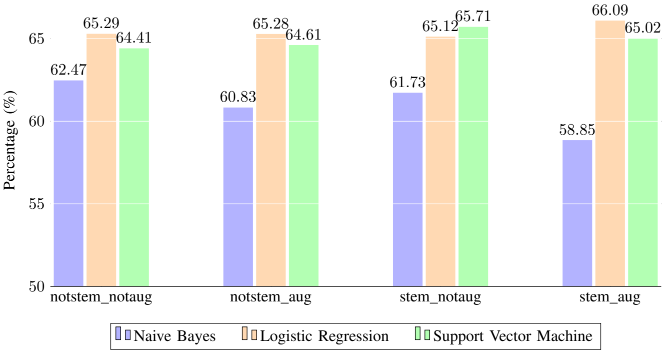


Fig. 8. The F_1 score results with TF-IDF configurations

(notstem: not-stemmed, stem: stemmed, notaug: not-augmented, aug: augmented)

The best configuration in the TF-IDF approach is "TFIDF_LR_stem_aug," i.e., the combination of logistic regression (LR), stemmed dataset (stem), and augmented dataset (aug). The model achieved 65.27% accuracy and 66.09% F_1 score. Furthermore, the worst configuration in the TF-IDF approach is from "TFIDF_NB_stem_aug" with 59.59% accuracy and 58.85% F_1 score. The experimental results show that different machine learning classifiers greatly impact the same preprocessed dataset, i.e., stemmed dataset (stem) and augmented dataset (aug). Moreover, the performance of the Naive Bayes classifier is inferior to the others.

From the machine learning classifier perspective, the logistic regression provides the best performance across all configurations, i.e., 64.42% average accuracy and 65.45% average F_1 score. From the data preprocessing perspective, using the stemming process increases the overall accuracy by 0.11% compared to not using it. However, it reduces the overall F_1 score by 0.06%. Finally, the augmentation process reduces average accuracy and average F_1 score by 0.45% and 0.67%, respectively.

Fig. 9 shows the confusion matrix result of the TF-IDF model, i.e., "TFIDF_LR_stem_aug". The most accurate prediction is anger (18.39%), and the least accurate prediction

[4] https://github.com/rhiosutoyo/emotion-recognition-model

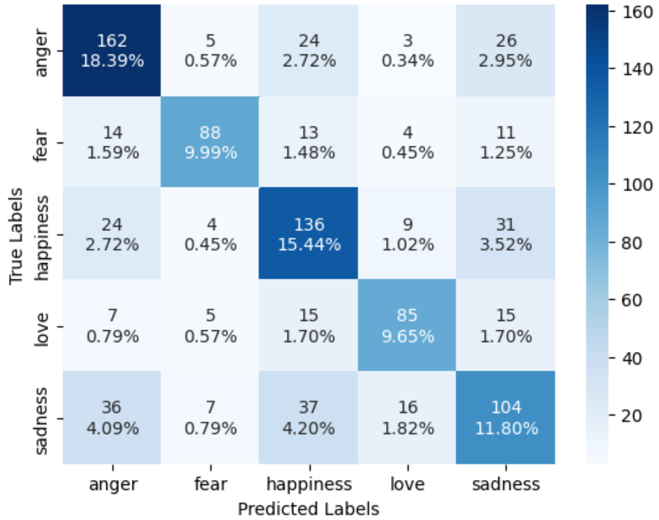


Fig. 9. The confusion matrix from the best TF-IDF model (TFIDF_LR_stem_aug).

is love (9.65%). The result is unsurprising because anger has the highest label than the other emotions. Furthermore, love has the lowest label than the other emotions. The confusion matrix result shows that anger tends to be misinterpreted as sadness. Moreover, fear tends to be misinterpreted as anger. Happiness tends to be misinterpreted as sadness. Love tends to be misinterpreted as happiness or sadness. Lastly, sadness tends to be misinterpreted as happiness.

Based on the experiment results, implementing data augmentation to increase the training data does not yield a positive outcome. In theory, the augmentation process is supposed to increase the model performance. However, the performance of using augmentation techniques is lower than that of not using them. This work argues that the original dataset has performed well because the Indonesian Twitter dataset is quite balanced. Thus, the augmented dataset is having trouble outperforming the performance of the not-augmented dataset.

C. Word Embedding

This work explored 27 feature engineering configurations of word embedding (WE) by combining three-word embedding models, three pooling techniques, and three machine-learning classifiers. The F_1 score results for every configuration with word embedding are presented in Fig. 10.

The best configuration in the word embedding approach is from the combination of support vector machine (SVM), fastText (ft), and sum pooling (sum), or “WE_SVM_ft_sum.” The model achieved 65.27% accuracy and 64.50% F_1 score. Furthermore, the worst configuration in the word embedding approach is from “WE_NB_w2v_sum” with 38.59% accuracy and 36.86% F_1 score.

From the machine learning classifier perspective, the SVM performs best across all configurations, i.e., 58.17% average accuracy and 57.89% average F_1 score. From the word embedding model’s perspective, the fastText model provides the best performance across all configurations, i.e., 56.26% average accuracy and 55.68% average F_1 score. The fastText’s average F_1 score is 3.52% higher than Google’s NNLM and 3.7% higher than the word2vec model. The word2vec incapability to handle out-of-vocabulary tokens might be the reason for the poor performance. Lastly, the pooling technique with the best average performance across all configurations is mean pooling, i.e., 57.28% average accuracy and 56.83% average F_1 score. The mean pooling’s average F_1 score is 7.12% higher than the min-max pool and 2.49% higher than the sum pool. The min-max pooling technique that discards features from the input might be the reason for the poor performance.

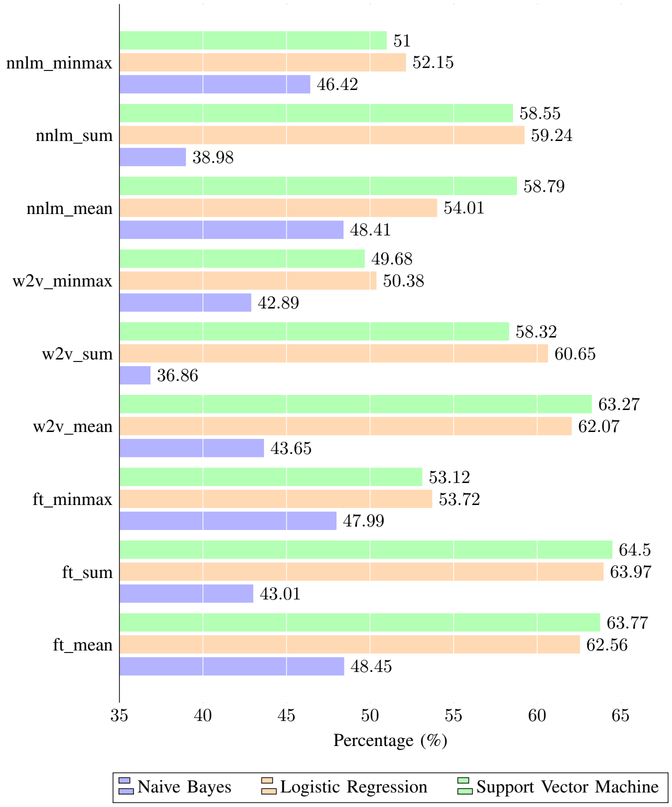


Fig. 10. The F_1 score results with word embedding configurations.

(nnlm: neural network language model, w2v: word2vec, ft: fastText)

Fig. 11 shows the confusion matrix result of the word embedding model, i.e., “WE_SVM_ft_sum”. The most accurate prediction is anger (21.11%), and the least accurate is sadness (8.29%). Emotion with the highest result from word embedding is the same as the TF-IDF approach,i.e., anger. Emotion with the lowest result from word embedding differs from the TF-IDF approach, i.e., love. The confusion matrix result shows that anger tends to be misinterpreted as happiness. Furthermore, fear tends to be misinterpreted as anger. Happiness tends to be misinterpreted as anger. Love tends to be misinterpreted as happiness. This is normal because both emotions share similar characteristics. Lastly, sadness tends to be misinterpreted as anger.

Unlike the TF-IDF, the confusion matrix result of word embedding shows that the prediction results do not align with the sum of the label quantities in the dataset. Based on the sum

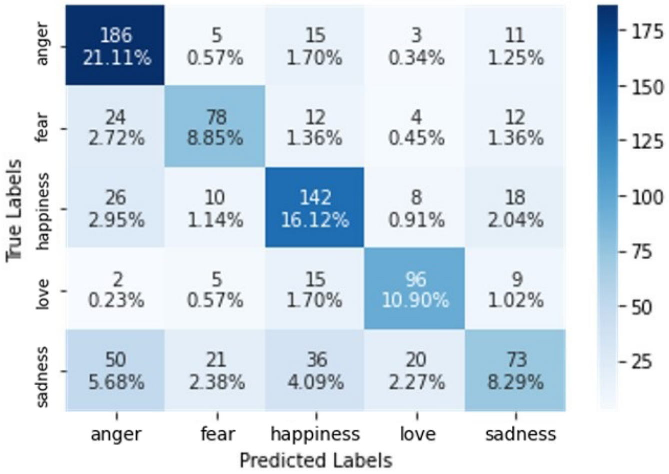


Fig. 11. The confusion matrix from the best word embedding model (WE_SVM_ft_sum).

TABLE II. The Configuration Detail of the Top Seven Models in TF-IDF

Configuration Component	Measurement	Frequency	%
Data Stemming	Stemmed	4	57%
	Not-Stemmed	3	43%
	Total	7	100%
Data Augmentation	Augmented	4	57%
	Not-Augmented	3	43%
	Total	7	100%
Machine Learning Classifier	Logistic Regression	4	57%
	Support Vector Machine	3	43%
	Naive Bayes	0	0%
	Total	7	100%

of the label quantities in the dataset, the order of emotions is anger, happiness, sadness, fear, and love (see Fig. 2). Based on the confusion matrix performance, the emotions are anger, happiness, love, fear, and sadness (see Fig. 11).

V. Discussions

A. Configuration Detail of the Top Seven Models

This work proposed 39 feature engineering configurations, i.e., 12 with TF-IDF and 27 with word embedding. This section shows the configuration detail of the top seven models from both word representation techniques.

The configuration detail of the top seven models in TF-IDF is shown in Table II. The experiment shows that the top-performance model has the following component configurations: logistic regression, stemmed dataset, and augmented dataset. The result matches the top model on TF-IDF, which is "TFIDF_LR_stem_aug."

Moreover, the configuration detail of the top seven models in word embedding is shown in Table III. The experiment shows that the top-performance model has the following component configurations: logistic regression, fastText, and mean pooling. The result does not match the top word embedding model, "WE_SVM_ft_sum."

TABLE III. The Configuration Detail of the Top Seven Models in Word Embedding

Configuration Component	Measurement	Frequency	%
Word Embedding Model	fastText	4	57%
	word2vec	3	43%
	NNLM	0	0%
	Total	7	100%
Pooling Technique	Mean Pooling	4	57%
	Sum Pooling	3	43%
	Min-Max Pooling	0	0%
	Total	7	100%
Machine Learning Classifier	Logistic Regression	4	57%
	Support Vector Machine	3	43%
	Naive Bayes	0	0%
	Total	7	100%

TABLE IV. Average F_1 Score of TF-IDF and Word Embedding from the Top Seven Model Configurations

Word Representation	F_1 Score
TF-IDF	65.30%
Word Embedding	62.97%

The top seven model configurations of TF-IDF and word embedding produce an F_1 score of more than 60%. On average, the top seven models of TF-IDF perform better than word embedding. The results are shown in Table IV.

B. TF-IDF vs. Word Embedding

Based on the experiment, the F_1 score from the best model of the word embedding technique (64.5%) is lower than the TF-IDF technique (66.09%). Hence, the best feature engineering configuration for the emotion recognition model is the TF-IDF approach from the combination of logistic regression (LR), stemmed dataset (stem), and augmented dataset (aug).

In theory, the word embedding technique should be able to produce a higher performance result because it has dense information packed in fixed-size arrays. Nevertheless, other research also shows that the performance of word embedding is lower than TF-IDF [23], [24]. In their research [23], Piskorski and Jacquet argue that features from word embedding might be great for deep learning but not for machine learning. Furthermore, specific features make the dataset biased in favor of traditional machine-learning approaches. These features appear exclusively for some categories (e.g., unique keywords). Thus, the classical machine learning algorithms can perform the classification with high precision because of the feature vector built using the TF-IDF technique.

C. Performance Comparison

The previous work from [13] does not provide codes and test splits. Hence, this study repartitioned the Indonesian Twitter emotion dataset by using several random seed values to perform the experiments. Ultimately, the random seed "88" was chosen because it achieved the best result.

The experiment resulted in a slightly better performance than the previous study [13] from the perspective of basic features, i.e., Word2Vec (WV) and FastText (FT). The Word2Vec (WV) F_1 score is increased by 1.44% by using mean pooling and SVM. Moreover, the FastText (FT) F_1 score is increased by 2.01% by using sum pooling and SVM.

In general, the quality of features can be increased by utilizing different pooling methods. Moreover, SVM produces better results than logistic regression.

VI. Conclusions and Future Work

This research explored feature engineering to build an emotion recognition model on the Indonesian Twitter dataset. Two different text data representations were used, namely TF-IDF and word embedding. This research proposes 12 feature engineering configurations in TF-IDF by mixing data stemming, data augmentation, and machine learning classifiers. Furthermore, this research proposes 27 feature engineering configurations in word embedding by mixing three-word embedding models, three pooling techniques, and three machine-learning classifiers. Moreover, this research analyzed the top seven models of both data representation techniques to find the recommended configuration component. Finally, performance comparisons were conducted to evaluate the models further.

The best performance configuration of TF-IDF is achieved by "TFIDF_LR_stem_aug," i.e., logistic regression (LR), stemmed dataset (stem), and augmented dataset (aug). The model achieved 65.27% accuracy and 66.09% F_1 score. The best performance configuration of word embedding is achieved by "WE_SVM_ft_sum," i.e., support vector machine (SVM), fastText (ft), and sum pooling (sum). The model achieved 65.27% accuracy and 64.50% F_1 score. The detailed characteristics from the top seven models show the recommended component configurations in TF-IDF: logistic regression, stemmed dataset, and augmented dataset. The recommended component configurations in word embedding are logistic regression, fastText, and mean pooling.

Furthermore, the experiment shows a slightly better performance than the previous study from the perspective of a single basic feature, i.e., word embedding. This work improved the word2vec F_1 score by 1.44% by using mean pooling and SVM. Moreover, the fastText F_1 score is increased by 2.01% by using sum pooling and SVM. Based on the results, the quality of text features in word embedding can be enhanced by utilizing different pooling methods. The recommended configuration element in word2vec is mean pooling; in fastText, it is sum pooling.

Lastly, the experiment shows that word embedding performs lower than TF-IDF. Thus, further exploration of utilizing word embedding in deep learning with a more significant number of examples can become the focus of future work.

References

[1] A. S. Girsang, S. M. Isa, and I. Harvy, "Recommendation System Journalist For Getting Top News Based On Twitter Data," *J. Phys. Conf. Ser.*, vol. 1807, no. 1, p. 012006, Apr. 2021.

[2] X. Guo and J. Li, "A novel twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, 2019, pp. 472–477.

[3] P. R. Shaver, U. Murdaya, and R. C. Fraley, "Structure of the indonesian emotion lexicon," *Asian journal of social psychology*, vol. 4, no. 3, pp. 201–224, 2001.

[4] G. Mohammadi and P. Vuilleumier, "A multi-componential approach to emotion recognition and the effect of personality," *IEEE Transactions on Affective Computing*, 2020.

[5] W. Budiharto and M. Meiliana, "Prediction and analysis of indonesia presidential election from twitter using sentiment analysis," *Journal of Big data*, vol. 5, no. 1, pp. 1–10, 2018.

[6] A. Andrew, A. N. Tjokrosetio, and A. Chowanda, "Dynamic difficulty adjustment with facial expression recognition for improving player satisfaction in a survival horror game," *ICIC Express Letters*, vol. 14, no. 11, pp. 1097–1104, 2020.

[7] B. Siswanto, F. L. Gaol, B. Soewito, and H. L. H. S. Warnars, "Sentiment analysis of big cities on the island of java in indonesia from twitter data as a recommender system," in *2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS*. IEEE, 2021, pp. 124–128.

[8] R. Sutoyo, H. L. H. S. Warnars, S. M. Isa, and W. Budiharto, "Emotionally aware chatbot for responding to indonesian product reviews," *International Journal of Innovative Computing, Information and Control*, vol. 19, no. 03, p. 861, 2023.

[9] F. R. Lapitan, R. T. Batista-Navarro, and E. Albacea, "Crowdsourcing-based annotation of emotions in filipino and english tweets," in *Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP2016)*, 2016, pp. 74–82.

[10] S. Mohammad and F. Bravo-Marquez, "WASSA-2017 shared task on emotion intensity," in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, A. Balahur, S. M. Mohammad, and E. van der Goot, Eds. Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017, pp. 34–49. [Online]. Available: https://aclanthology.org/W17-5205

[11] N. A. S. Winarsih, C. Supriyanto *et al.*, "Evaluation of classification methods for indonesian text emotion detection," in *2016 International seminar on application for technology of information and communication (ISemantic)*. IEEE, 2016, pp. 130–133.

[12] K. S. Nugroho and F. A. Bachtiar, "Text-based emotion recognition in indonesian tweet using bert," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE, 2021, pp. 570–574.

[13] M. S. Saputri, R. Mahendra, and M. Adriani, "Emotion classification on indonesian twitter dataset," in *2018 International Conference on Asian Language Processing (IALP)*. IEEE, 2018, pp. 90–95.

[14] N. Rosli, N. Rajaee, and D. Bong, "Renica based music source separation for automatic music emotion classification," *INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING INFORMATION AND CONTROL*, vol. 14, no. 6, pp. 2325–2333, 2018.

[15] D. Wu, H. Zhao, X. Zhang, Z. Tao, and C. Huang, "Multi-scaled emotional recognition from speech using fuzzy model and markov random fields based configuration," *ICIC Express Lett.*, vol. 9, no. 6, pp. 1637–1642, Jan. 2015.

[16] H. Jung, M. Kwon, and H.-I. Cheng, "Analysis of EEG for violent movies," *ICIC Express Lett.*, vol. 10, no. 7, pp. 1523–1528, Jul. 2016.

[17] E. S. Dan-Glauser and K. R. Scherer, "The difficulties in emotion regulation scale (ders)," *Swiss Journal of Psychology*, 2012.

[18] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, 2005, pp. 579–586.

[19] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, 2007, pp. 70–74.

[20] D. A. Sauter, F. Eisner, P. Ekman, and S. K. Scott, "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations," *Proceedings of the National Academy of Sciences*, vol. 107, no. 6, pp. 2408–2412, 2010.

[21] M. Adriani, J. Asian, B. Nazief, S. M. Tahaghoghi, and H. E. Williams, "Stemming indonesian: A confix-stripping approach," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 6, no. 4, pp. 1–33, 2007.

[22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[23] J. Piskorski and G. Jacquet, "Tf-idf character n-grams versus word embedding-based models for fine-grained event classification: a preliminary study," in *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020*, 2020, pp. 26–34.

[24] A. W. Romadon, K. M. Lhaksmana, I. Kurniawan, and D. Richasdy, "Analyzing tf-idf and word embedding for implementing automation in job interview grading," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 2020, pp. 1–4.

Robust Extreme Learning Machine with Exponential Squared Loss via DC Programming

Kuaini Wang[1], Xiaoxue Wang[2], Weicheng Zhan[3], Mingming Wang[4], Jinde Cao[5]
Southeast University, School of Mathematics, China[1,5]
Xi'an Shiyou University, College of Science, China[1]
Xi'an Shiyou University, School of Computer Science, China[2,3,4]
Yonsei University, Yonsei Frontier Lab, South Korea[5]

***Abstract*—Extreme learning machines (ELM) have recently attracted considerable attention because of its fast learning rate, simple model structure, and good generalization ability. However, classical ELM with least squares loss function is prone to overfitting and lack robustness in dealing with datasets containing noise and outliers in the real world. In this paper, inspired by the maximum correntropy criterion, an exponential squared loss function is introduced, which is nonconvex and insensitive to noise and outliers. A robust ELM with exponential squared loss (RESELM) is presented to overcome the overfitting problem. The proposed model with nonconvexity is difficult to be directly optimized. Considering the superior performance of difference of convex functions (DC) programming in solving nonconvex problems, this paper optimizes the model by expressing the objective function as a DC function and employing DC algorithm (DCA). To examine the effectiveness of the proposed algorithm in noisy environment, different levels of outliers are added to the training samples in the experiments. Experimental results on benchmark data sets with different outliers levels illustrate that the proposed RESELM achieves significant advantages in generalization performance and robustness, especially in higher outliers levels.**

Keywords*—Extreme learning machine; exponential squared loss; DC programming; DCA; robust regression

I. Introduction

To improve the slow learning rate of single hidden layer feedforward neural networks (SLFNs), Huang and his team proposed extreme learning machine (ELM) in 2004 [1], [2], [3]. Traditional algorithms for feedforward neural networks require iterative adjustment of all parameters in the whole network. However, the experiments in [4], [5] illustrate that input weights and hidden layer bias of SLFNs may not need to be adjusted. The hidden layer input weights and biases of ELM are determined by random generation, which reduces the number of parameters to be solved in the network by a large part. ELM can be regarded as a simple linear system with only the output weights to be solved. ELM is widely used in various real-world problems relying on its fast learning rate, simple model structure, and good generalization ability [6], [7], [8].

However, samples in real-world problems are different from the clean and uncontaminated samples used in the laboratory, which is potentially polluted in the process of both generation and acquisition [9]. Training ELM with samples containing outliers can exacerbate the discrepancy between the true and predicted values, leading to longer learning time and poorer model prediction accuracy [10], [11]. The loss function plays a crucial role in ELM training. Classical ELM employs the least squares loss function, which is easy to be solved and can improve the learning rate of the model. However, its squared effect leads to more sensitivity to outliers. When the outliers are larger, the empirical risk of the model becomes higher, which eventually affects the accuracy of the model [12].

In order to minimize the disturbance of outliers, researchers have turned to finding alternative loss functions to obtain a more robust algorithm [13], [14]. Deng et al. proposed an improved ELM based on a weighted 2-norm loss function (WELM) [15], which assigned weights to the samples depending on the residuals, improved the model's generalization. Zhang et al. developed outlier robust ELM (ORELM) by applying the 1-norm loss function to ELM [16]. The 1-norm loss function grows slower than the 2-norm loss function as the residuals increase, thus obtaining a better accuracy than ELM. Chen et al. constructed a robust ELM that can use four loss functions (1-norm, Huber, Bisquare, Welsch) [17]. The experiment's optimal accuracy was obtained by the model that used Bisquare or Welsch loss functions, both of which are nonconvex loss functions. The 1-norm and Huber loss functions are both convex loss functions and has a linear relationship with residuals, which is still not robust. When the residuals are enormous, the penalty imposed on the sample by the convex loss function can also be very large. Models usually treat outliers as normal values to reduce the large value's loss caused by outliers in empirical risk at the cost of sacrificing the model's generalization, and nonconvex loss functions can compensate for this deficiency [18].

The nonconvex loss function has a strong learning capability in terms of both generalization and robustness. The capped type of nonconvex loss functions can directly limit the maximum penalty value caused by noise and outliers and explicitly suppress the negative impact of such samples on the decision hyperplane to build models with excellent robustness [18], [19]. Different capped 2-norm loss functions are constructed in [20] and [21], respectively, which have shown stronger robustness and generalization. In recent years, with the development of information theory, Liu et al. proposed correntropy in 2007 [22], which is a measure of similarity of two sets of random variables and widely used in robust learning. Xing et al. proposed a robust ELM model based on the regularized correntropy criterion [23], which showed that the proposed model has better robustness and can effectively handle scenes with outlier interference.

This paper proposes a nonconvex exponential squared loss function inspired by the above literature and the maximum correntropy criterion. This loss function is applied to ELM, which leads to a new robust ELM. RESELM can sufficiently suppress the negative impact of outliers on the robustness of the model and effectively improve the model's generalization. However, nonconvexity makes the model difficult to optimize. Considering the advantages of DCA [24] in solving nonconvex problems, this paper converts the objective function into DC programming [25] and then uses DCA to obtain the optimal output nodes.

The main contributions of this paper can be summarized as follows:

(1) A new loss function is constructed based on the maximum correntropy criterion, called the exponential squared loss function, which is nonconvex and can deal with training samples with noise and outliers.

(2) Robust ELM with the exponential squared loss function (RESELM) is developed. The nonconvexity of the proposed model makes it difficult to optimize directly by classical convex optimization algorithms. Therefore, it is transformed into a DC programming, and solved by DCA.

(3) The RESELM is tested in the case with 0%-20% lower outliers levels and with 25%-40% higher outliers levels, respectively. The experimental results show that RESELM improves the robustness and has excellent generalization ability, especially in the case of higher outliers levels.

The remainder of the paper is organized as follows. Section II briefly reviews the ELM. In Section III, the proposed model of this paper and the process of optimization by adopting DC programming are elaborated in detail. In Section IV, the experimental results of RESELM with different outliers levels are shown and analyzed. In the fifth part, the work of this paper is summarized.

II. Related Works

For N arbitrary samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, where $\mathbf{x}_i \in R^d$ is the input variable and $y_i \in R$ is the corresponding target in regression estimation, the output of ELM with L hidden nodes can be described as follows:

$$f(\mathbf{x}) = \sum_{j=1}^{L} h_j(\mathbf{x})\boldsymbol{\beta_j} = \mathbf{h}(\mathbf{x})\boldsymbol{\beta} \tag{1}$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_L]^T$ is the output weights vector that connects the hidden layer to the output node, and $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), h_2(\mathbf{x}), ..., h_L(\mathbf{x})]$ is the output of the hidden layer and $h_j(\mathbf{x})$ is the activation function. The formulation of regularized ELM [26] can be expressed as the following optimization:

$$\underset{\boldsymbol{\beta}}{min} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\sum_{i=1}^{N} e_i^2 \tag{2}$$

$$s.t. \quad \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = y_i - e_i, i = 1, ..., N, \tag{3}$$

where e_i denotes the error of training sample $\mathbf{x}_i$, and C is the regularization parameter. The optimal solution $\boldsymbol{\beta}$ of (2)-(3) is given by [26]

$$\boldsymbol{\beta} = \begin{cases} (\mathbf{H}^T\mathbf{H} + \dfrac{\mathbf{I}}{C})^{-1}\mathbf{H}^T\mathbf{y}, & N \geq L, \\ \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \dfrac{\mathbf{I}}{C})^{-1}\mathbf{y}, & N < L. \end{cases} \tag{4}$$

where $\mathbf{y} = [y_1, y_2, ..., y_N]^T$ and

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \mathbf{h}(\mathbf{x}_2) \\ . \\ . \\ . \\ \mathbf{h}(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{x}_1) & h_2(\mathbf{x}_1) & ... & h_L(\mathbf{x}_1) \\ h_1(\mathbf{x}_2) & h_2(\mathbf{x}_2) & ... & h_L(\mathbf{x}_2) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ h_1(\mathbf{x}_N) & h_2(\mathbf{x}_N) & ... & h_L(\mathbf{x}_N) \end{bmatrix} \tag{5}$$

is the output matrix of the hidden layer.

In Regularized ELM (2)-(3), the least squares loss function may result in poor performance of ELM in dealing with the training samples containing noise and outliers. The reason for this is that least squares loss function assumes that the training samples obey a normal error distribution [12]. However, it can not guaranteed in the real world, which mistakenly considers the role of outliers with large residuals. This paper will focus on suppressing the effect of outliers by introducing a new exponential squared loss function for the ELM.

III. Research Methodology

A. Exponential Squared Loss Function

In information theory, the maximum correntropy criterion [22] is used to deal with the analysis of signals affected by various noises, which can effectively improve the robustness of signal analysis, and it is defined as follows:

$$V_\sigma(A, B) = E[k_\sigma(A - B)], \tag{6}$$

where $k_\sigma(\cdot)$ is a kernel function and $E[\cdot]$ is the mathematic expectation. In general, the joint probability distribution between variables A and B is unknown, so the average value is used to estimate the mathematical expectation. Then the maximum correntropy criterion is expressed as

$$V_\sigma(A, B) = \frac{1}{m}\sum_{i=1}^{m} k_\sigma(A_i, B_i), \tag{7}$$

where $k_\sigma(A_i, B_i) = \exp\left(-\frac{\|A_i - B_i\|^2}{\sigma^2}\right)$ is Gaussian kernel function.

In order to overcome the drawback of the least squares loss function, this paper constructs the exponential squared loss function based on the correntropy,

$$\ell_\sigma(z) = \sigma^2\left[1 - \exp\left(-\frac{z^2}{\sigma^2}\right)\right], \tag{8}$$

where σ^2 is the upper bound of the exponential squared loss function. Fig. 1 demonstrates the different curves of exponential squared loss with respect to the different of σ^2. As shown in Fig. 1, the proposed loss function is bounded.

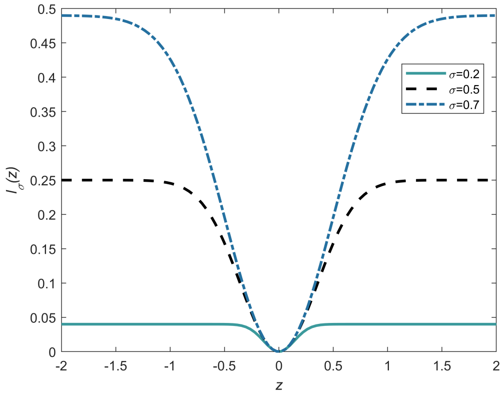


Fig. 1. Exponential squared loss function with different σ.

B. Robust ELM with Exponential Squared Loss Function

In this subsection, a robust ELM with exponential squared loss function is developed to improve the robustness of ELM, and the corresponding optimization problem can be obtained as

$$\min_{\boldsymbol{\beta}} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\sum_{i=1}^{N} \ell_\sigma(z_i) \tag{9}$$

where training error $z_i = y_i - \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta}$ and $\ell_\sigma(z_i)$ is the exponential squared loss function. The expression (8) can be written in the following equivalent form.

$$\ell_\sigma(z) = \ell_1(z) - \ell_2(z) \tag{10}$$

where $\ell_1(z) = z^2$, $\ell_2(z) = z^2 - \sigma^2\left[1 - \exp\left(-\frac{z^2}{\sigma^2}\right)\right]$. Substituting $\ell_1(z)$ and $\ell_2(z)$ into the optimization problem (9) can be derived as follows:

$$\min_{\boldsymbol{\beta}} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\sum_{i=1}^{N} \ell_1(z_i) - \frac{C}{2}\sum_{i=1}^{N} \ell_2(z_i) \tag{11}$$

Mark $L_1(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{\beta}\| + \frac{C}{2}\sum_{i=1}^{N}\ell_1(z_i)$, $L_2(\boldsymbol{\beta}) = \frac{C}{2}\sum_{i=1}^{N}\ell_2(z_i)$. According to the DCA, the optimal solution of the optimization problem (11) is obtained by solving the following iterations:

$$\boldsymbol{\beta}^{(t+1)} = \arg\min_{\boldsymbol{\beta}} \left\{ L_1(\boldsymbol{\beta}) - L_2^{'}(\boldsymbol{\beta}^{(t)}) \cdot \boldsymbol{\beta} \right\} \tag{12}$$

where $L_2^{'}(\boldsymbol{\beta}^{(t)})$ represents the derivative of $L_2(\boldsymbol{\beta})$ at $\boldsymbol{\beta}^{(t)}$. For a certain $\boldsymbol{\beta}$, the derivative expression is as follows:

$$L_2^{'}(\boldsymbol{\beta}) = \frac{C}{2}\sum_{i=1}^{N} \frac{\partial \ell_2(z_i)}{\partial z_i} \cdot \frac{\partial z_i}{\partial \boldsymbol{\beta}} = \frac{C}{2}\sum_{i=1}^{N} \left(\frac{\partial \ell_2(z_i)}{\partial z_i}\right) \cdot \left(-\mathbf{h}^T(\mathbf{x}_i)\right) \tag{13}$$

Denote $s_i = \frac{C}{2} \cdot \frac{\partial \ell_2(z_i)}{\partial z_i}$, and then define

$$s_i = C z_i \left[1 - \exp\left(-\frac{z_i^2}{\sigma^2}\right)\right] \tag{14}$$

From (13) and (14),

$$\begin{aligned} -L_2^{'}\left(\boldsymbol{\beta}^{(t)}\right) \cdot \boldsymbol{\beta} &= \sum_{i=1}^{N} s_i^{(t)} \cdot \mathbf{h}\left(\mathbf{x}_i\right)\boldsymbol{\beta} \\ &= \mathbf{s}^T \mathbf{H}\boldsymbol{\beta} \end{aligned} \tag{15}$$

where $\mathbf{s} = [s_1, s_2, ..., s_N]^T$, then (12) can be transformed into solving the following optimization problem

$$\boldsymbol{\beta}^{(t+1)} = \arg\min \left\{ \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\|\mathbf{y} - \mathbf{H}\boldsymbol{\beta}\|^2 + \mathbf{s}^T\mathbf{H}\boldsymbol{\beta} \right\} \tag{16}$$

Following the line [26], the optimal solution of (12) in the (t+1) iteration is obtained

$$\boldsymbol{\beta} = \begin{cases} \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1} \mathbf{H}^T \left(\mathbf{y} - \frac{\mathbf{s}}{C}\right) & N > L, \\ \mathbf{H}^T \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1} \left(\mathbf{y} - \frac{\mathbf{s}}{C}\right) & N \le L. \end{cases} \tag{17}$$

Next is the step to solve RESELM by DC algorithm

Algorithm 1 RESELM

```
Input: $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, set $t$=0 and choose an initial point $\mathbf{s}^{(0)}$,
  $L$, $C$, $t_{max}$ and $\varepsilon > 0$ is a sufficient small number
Output: $\boldsymbol{\beta}$
  repeat
    Compute $s_i^{(t)}$ by (14) and $\mathbf{s}$.
    Calculate (17) to obtain $\boldsymbol{\beta}^{(t+1)}$.
    Let $t$=$t$+1.
  until $\left\|\boldsymbol{\beta}^{(t)} - \boldsymbol{\beta}^{(t+1)}\right\| \le \varepsilon$ or $t > t_{max}$
```

IV. Results and Discussion

To validate the efficacy of RESELM, it is compared with four related methods, regularized ELM [26], weighted ELM (WELM) [15], outlier robust ELM (ORELM) [16] and iteratively reweighted ELM (IRWELM) [27] on 18 benchmark data sets. In the first part, to simulate the samples in real world, different outliers levels are added to each of the 18 benchmark data sets, which can better reflect the robustness. of each algorithm. In the second part, the effects of the number of hidden nodes L and the parameter σ on the performance of the algorithms are studied. The root mean squares error (RMSE) [28] is used to measure the performance of the five regression algorithms. In the experiments, three parameters should be selected, L, C, σ. The maximum number of iteration t_{max} = 20, and the number of hidden nodes L=200. C is chosen from $\left\{2^{-19}, 2^{-18}, 2^{-17}, ..., 2^{18}, 2^{19}, 2^{20}\right\}$, and the width parameter of exponential squared loss function in RESELM is chosen from $\{0.05, 0.1, 0.15, 0.2, ..., 0.95, 1\}$.

A. Experimental Results on Benchmark Data Sets

This section shows the accuracy of the five algorithms on the benchmark data sets with different outliers levels. Table 1 demonstrates the algorithms' RMSE on 10 benchmark data sets with different outliers levels (0%, 5%, ...,35%, 40%). Fig. 2 adopts a line chart to intuitively exhibit the accuracy variation

TABLE I. EXPERIMENTAL RESULTS ON BENCHMARK DATA SETS WITH DIFFERENT OUTLIERS LEVELS

	Outliers	ELM	WELM	ORELM	IRWELM	RESELM
Data set	levels	(RMSE ± Std)	(RMSE ± Std)	(RMSE ± Std)	(RMSE ± Std)	(RMSE ± Std)
Diabetes	0%	0.5832±0.0935	**0.5818±0.0908**	0.5946±0.1012	**0.5818±0.0907**	0.5819±0.0926
	5%	0.6480±0.0995	0.5757±0.0944	0.5985±0.0726	**0.5734± 0.0898**	0.5740±0.0906
	10%	0.6486±0.1028	0.5897±0.1036	0.6151±0.0709	**0.5781±0.0982**	0.5787±0.0828
	15%	0.6658±0.1131	0.6091±0.1140	0.6342±0.1031	0.5853±0.0946	**0.5745±0.0837**
	20%	0.6625±0.1232	0.6514±0.0934	0.6314±0.1057	0.5967±0.0925	**0.5795±0.0953**
	25%	0.6978±0.1244	0.7372±0.1979	0.6346±0.1081	0.7666±0.1686	**0.5801±0.0982**
	30%	0.6732±0.1243	0.7437±0.1923	0.6445±0.1097	0.7570±0.2105	**0.5823±0.0941**
	35%	0.6641±0.1230	0.6641±0.1230	0.6638±0.1022	0.6641±0.1230	**0.6036±0.0963**
	40%	0.6623±0.1184	0.6647±0.1192	0.6601±0.1051	0.6655±0.1196	**0.6224±0.1120**
Pollution	0%	36.0203±6.6156	37.0793±4.1013	36.2977±5.6007	37.2221±4.0031	**35.9914±6.4602**
	5%	56.7646±6.9775	37.7693±6.2724	37.6903±6.1268	37.2284±5.8706	**36.1362±6.7496**
	10%	57.4947±7.8601	39.5232±6.5899	39.5366±6.3062	38.1760±5.7391	**36.6225±5.9713**
	15%	59.5536±9.4949	43.3163±4.5347	43.0358±6.8460	38.3071±6.1082	**36.9604±5.3304**
	20%	59.3307±7.8055	48.0174±11.8658	45.0034±7.3071	37.9233±6.2045	**37.5655±5.9203**
	25%	64.6370±12.6879	58.7079±9.5638	49.6946±9.4306	65.1266±12.7976	**38.6563±6.0681**
	30%	61.5882±11.2617	61.5882±11.2617	56.4876±13.6835	61.5882±11.2617	**39.2598±5.7079**
	35%	58.8763±9.0385	58.8763±9.0385	57.8867±8.6369	58.8763±9.0385	**39.7484±6.1533**
	40%	59.8161±8.7125	59.8161±8.7125	57.8640±7.6647	59.8161±8.7125	**40.4083±5.1636**
Pyrim	0%	0.1114±0.0204	0.1060±0.0297	0.1084±0.0256	0.1077±0.0291	**0.1052±0.0318**
	5%	0.1305±0.0276	0.1083±0.0327	0.1098±0.0277	0.1078±0.0317	**0.1049±0.0317**
	10%	0.1315±0.0248	0.1127±0.0346	0.1134±0.0293	0.1103±0.0335	**0.1099±0.0352**
	15%	0.1402±0.0283	0.1199±0.0344	0.1182±0.0318	0.1092±0.0339	**0.1091±0.0342**
	20%	0.1451±0.0208	0.1289±0.0264	0.1207±0.0320	0.1107±0.0351	**0.1104±0.0339**
	25%	0.1469±0.0257	0.1405±0.0239	0.1284±0.0371	0.1190±0.0358	**0.1135±0.0322**
	30%	0.1456±0.0199	0.1381±0.0315	0.1313±0.0400	0.1293±0.0369	**0.1178±0.0325**
	35%	0.1549±0.0396	0.1556±0.0341	0.1389±0.0304	0.1559±0.0330	**0.1199±0.0363**
	40%	0.1583±0.0217	0.1587±0.0429	0.1396±0.0199	0.1580±0.0427	**0.1241±0.0356**
Servo	0%	0.6177±0.1013	**0.5547±0.1686**	0.6007±0.1416	0.6151±0.1666	0.5881±0.2043
	5%	0.8056±0.1097	0.7016±0.2281	0.6612±0.1732	0.7022±0.2212	**0.6455±0.1954**
	10%	0.9855±0.1307	0.7931±0.2099	0.7144±0.1969	0.7172±0.1868	**0.6827±0.1888**
	15%	1.0425±0.1378	0.8122±0.1500	0.7128±0.1816	**0.6935±0.1978**	0.6992±0.2195
	20%	1.0965±0.1633	0.9197±0.1655	0.7981±0.1871	0.7821±0.2080	**0.7521±0.1905**
	25%	1.3023±0.1662	0.9976±0.1218	0.7994±0.2028	0.7898±0.1995	**0.7341±0.2163**
	30%	1.4394±0.1417	1.1480±0.1802	0.9342±0.1977	0.9709±0.1605	**0.8323±0.2103**
	35%	1.4957±0.1387	1.4858±0.1487	1.1048±0.1575	1.4600±0.2192	**0.8663±0.2129**
	40%	1.4938±0.1681	1.4999±0.1360	1.2504±0.1558	1.4996±0.1361	**0.9211±0.1720**
Triazines	0%	0.1478±0.0169	0.1494±0.0189	0.1508±0.0203	0.1509±0.0198	**0.1471±0.0183**
	5%	0.1525±0.0180	0.1491±0.0197	0.1518±0.0200	0.1502±0.0196	**0.1487±0.0191**
	10%	0.1593±0.0153	0.1502±0.0212	0.1533±0.0205	0.1502±0.0199	**0.1496±0.0204**
	15%	0.1598±0.0170	0.1513±0.0192	0.1554±0.0206	0.1517±0.0200	**0.1507±0.0197**
	20%	0.1612±0.0138	0.1588±0.0157	0.1578±0.0160	0.1546±0.0154	**0.1527±0.0183**
	25%	0.1609±0.0140	0.1593±0.0165	0.1580±0.0164	0.1578±0.0200	**0.1574±0.0215**
	30%	0.1622±0.0177	0.1602±0.0173	0.1596±0.0172	0.1599±0.0162	**0.1577±0.0196**
	35%	0.1652±0.0122	0.1646±0.0196	0.1597±0.0157	0.1643±0.0188	**0.1583±0.0206**
	40%	0.1652±0.0122	0.1646±0.0186	**0.1595±0.0167**	0.1636±0.0176	0.1600±0.0154
MCPU	0%	54.0322±21.9741	59.0387±25.1172	**51.6258±25.1990**	54.8509±24.0867	51.7506±21.6742
	5%	79.5067±24.9900	55.1369±22.8733	58.2620±23.6020	**53.3418±23.6794**	55.1113±26.0012
	10%	114.6473±7.7572	61.2738±23.1768	58.5051±21.3603	57.3894±26.6030	**53.2878±23.2898**
	15%	116.6586±11.1286	70.2727±24.2272	63.9712±20.8597	60.6010±26.4665	**60.4078±25.1280**
	20%	145.8573±9.8285	78.6395±13.3020	72.2602±17.3181	**68.3295±26.7257**	70.3807±26.6680
	25%	158.6027±46.1825	84.5457±9.5708	73.1470±15.5764	78.3080±24.5267	**70.9848±22.4826**
	30%	159.1254±44.2106	96.0923±19.2900	72.4842±17.2320	86.3810±19.4879	**66.6127±25.0384**
	35%	159.8816±42.8364	143.0014±24.4416	79.7272±17.8036	124.9419±26.7260	**73.9987±18.2163**
	40%	158.9709±48.5852	160.5926±41.8149	88.6661±16.9791	160.4793±41.9984	**86.0494±25.0028**
Bodyfat	0%	0.0027±0.0015	0.0022±0.0017	**0.0021±0.0018**	**0.0021±0.0018**	0.0022±0.0018
	5%	0.0210±0.0018	0.0027±0.0015	**0.0022±0.0018**	**0.0022±0.0018**	0.0026±0.0016
	10%	0.0221±0.0026	0.0028±0.0015	**0.0022±0.0018**	**0.0022±0.0018**	0.0026±0.0016
	15%	0.0197±0.0021	0.0034±0.0014	**0.0022±0.0018**	**0.0022±0.0018**	0.0027±0.0015
	20%	0.0227±0.0033	0.0052±0.0018	**0.0022±0.0017**	**0.0022±0.0018**	0.0028±0.0015
	25%	0.0201±0.0011	0.0233±0.0056	**0.0023±0.0018**	0.0215±0.0040	0.0028±0.0015
	30%	0.0204±0.0025	0.0204±0.0025	**0.0024±0.0018**	0.0204±0.0025	0.0028±0.0015
	35%	0.0234±0.0048	0.0234±0.0048	0.0032±0.0018	0.0234±0.0048	**0.0030±0.0014**
	40%	0.0232±0.0039	0.0232±0.0039	0.0069±0.0024	0.0232±0.0039	**0.0034±0.0013**
AutoMPG	0%	2.8927±0.1359	**2.8697±0.1619**	2.9273±0.1445	2.9391±0.1228	2.8811±0.1753
	5%	3.4544±0.1374	2.9082±0.1223	2.9241±0.1439	2.9038±0.1366	**2.8781±0.1393**
	10%	4.1467±0.2067	2.9201±0.1447	2.9290±0.1170	2.8888±0.0989	**2.8869±0.0785**
	15%	5.2720±0.3020	3.0450±0.2378	2.9650±0.1801	**2.8866±0.1050**	2.8976±0.0992
	20%	5.9322±0.3485	3.3159±0.3372	2.9931±0.1597	2.9208±0.1188	**2.8964±0.1243**
	25%	7.6352±0.3434	5.1168±0.5234	3.1203±0.2035	4.0939±0.8859	**2.9122±0.1548**
	30%	7.9646±0.2458	7.8468±0.7013	3.3235±0.2356	7.1337±0.4989	**2.9339±0.1536**
	35%	8.0838±0.2344	8.0070±0.2422	3.8620±0.5010	8.0042±0.2448	**3.0375±0.2151**
	40%	8.2152±0.2736	8.1623±0.2287	4.9058±0.4914	8.1617±0.2289	**3.1374±0.1495**

Data set	Outliers levels	ELM (RMSE ± Std)	WELM (RMSE ± Std)	ORELM (RMSE ± Std)	IRWELM (RMSE ± Std)	RESELM (RMSE ± Std)
BH	0%	3.3436±0.2752	3.4172±0.4053	3.4892±0.4099	3.5044±0.4771	**3.3299±0.2751**
	5%	4.4329±0.5026	3.5431±0.4199	3.6025±0.4778	3.6465±0.3939	**3.4886±0.3914**
	10%	5.3008±0.4367	3.7162±0.7069	3.7211±0.6146	3.6627±0.5360	**3.5933±0.5327**
	15%	6.4081±0.3513	3.9997±0.4288	3.8031±0.4840	3.6272±0.3429	**3.6239±0.4093**
	20%	7.4551±0.3277	4.5851±0.5451	4.0599±0.4876	3.9450±0.5606	**3.8697±0.5311**
	25%	8.7010±0.2298	6.0587±0.4800	4.4545±0.6318	4.8437±0.5504	**4.1625±0.5608**
	30%	9.1915±0.5201	8.1431±0.6531	4.9708±0.6748	7.2347±0.6227	**4.3934±0.8690**
	35%	9.0939±0.4877	9.2483±0.4948	5.7650±0.4941	9.2744±0.4934	**4.7559±0.9979**
	40%	9.0767±0.4749	9.1630±0.5199	6.7831±0.6937	9.1791±0.5311	**5.2372±0.6306**
Concrete	0%	**6.6012±0.3947**	6.7518±0.4287	6.9974±0.5071	6.7847±0.4490	6.6016±0.3934
	5%	8.5968±0.4431	7.1756±1.0593	7.3170±0.6440	7.0391±0.8160	**6.7927±0.8174**
	10%	9.8699±0.5149	7.8765±0.4552	7.9495±0.5064	7.6314±0.4311	**7.3815±0.3971**
	15%	11.6955±0.5531	8.2027±0.3171	8.3026±0.4968	7.8751±0.3722	**7.7131±0.4287**
	20%	11.8756±0.5302	8.5947±0.2092	8.6132±0.4671	8.0975±0.3241	**7.9589±0.6622**
	25%	13.3383±0.5713	9.4221±0.2403	9.2141±0.5663	8.6074±0.2260	**8.3641±0.3404**
	30%	16.4737±0.6057	12.3405±0.8420	10.1421±0.7395	10.5091±0.9161	**8.4881±0.4038**
	35%	17.2334±0.3850	16.9023±0.9519	10.9747±0.6482	15.6942±1.0658	**8.9344±0.5123**
	40%	17.1472±0.3705	17.1077±0.3583	12.0187±0.5392	17.0847±0.3563	**9.0388±0.4906**

of the five algorithms on 8 benchmark data sets with 0% to 40% outliers levels.

In the experiments, each data set is randomly divided into training set and test set. Then different proportions of outliers are added to the training set, which is determined by the targets of the training set. The experiment select different proportions of outliers from $[y_{min},\ y_{max}]$ and add these outliers to the training samples randomly. The test set does not take any operation. The 10-fold cross-validation is applied on benchmark data sets, and taking the average RMSE of these ten independent experiments as the final result.

Observing the results in Table I, RESELM performs the best in the case of no outliers and achieves the optimal RMSE on four data sets. The accuracy on the other data sets is similar to the optimal RMSE. The worst performer is ELM, which achieves the optimum only on the Concrete data set. The performance of ORELM and IRWELM is close to each other. On the data set with 5% outliers level, the accuracy of ELM decreases most significantly, with a larger RMSE than the other algorithms on each data set. The accuracy of IRWELM is better than it would have been in the case of no outliers. RESELM obtains more comparable robustness, and its accuracy is optimal on most of the data sets. When the outliers level rises to 10%, the accuracy of ELM is still not very competitive, and the RMSE of WELM is better than that of the ELM. RESELM maintains its advantage in robustness, obtaining the optimal RMSE on eight data sets. The accuracy of the IRWELM is next to that of the RESELM in most cases. In the case of 15% and 20% outliers levels, the accuracy of RESELM obtains the best RMSE on most of the data sets. At the same time, ELM fails to obtain the optimal accuracy on any of data sets and has the worst RMSE of the five algorithms in most cases.

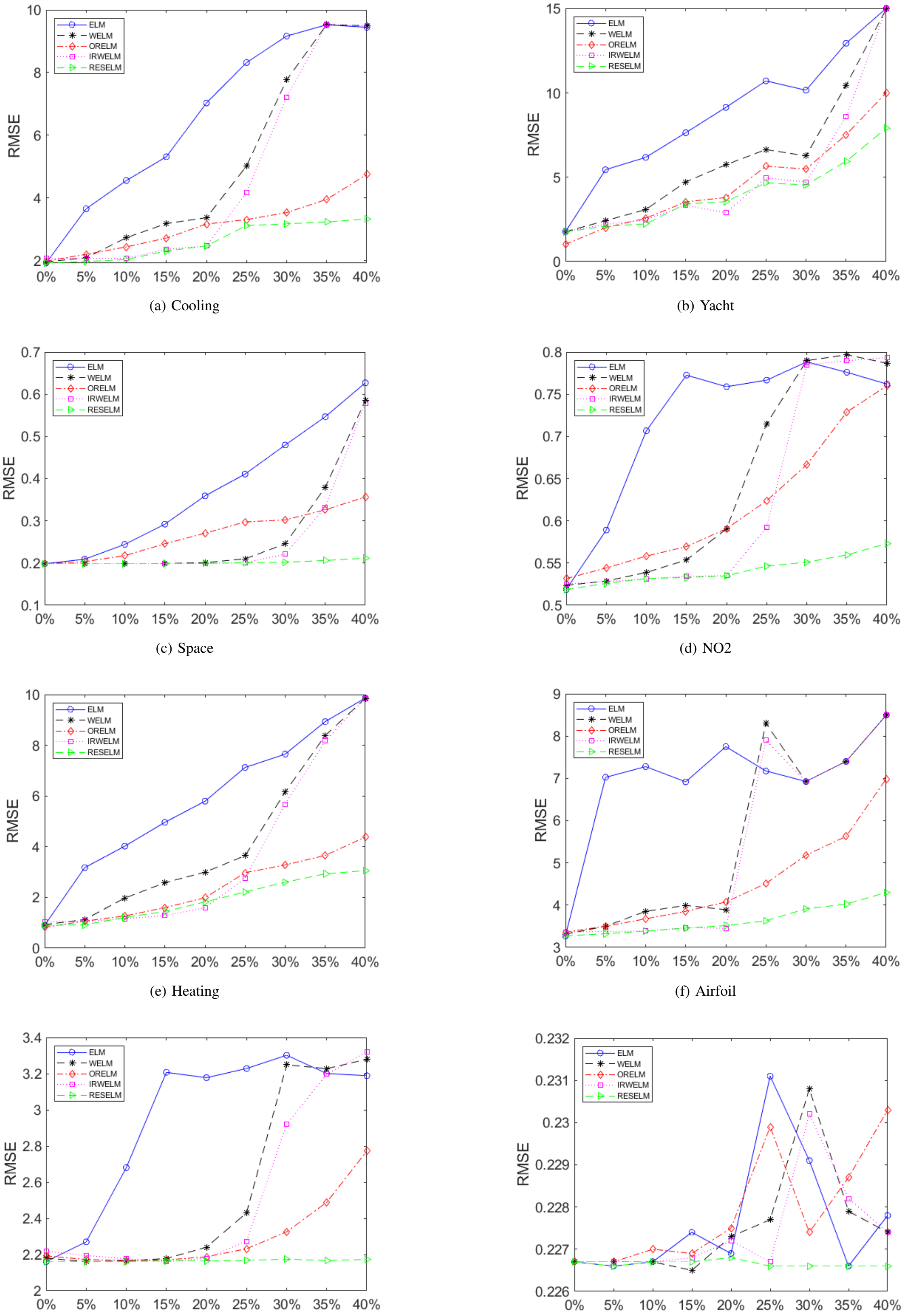
In the lower outliers levels (0%-20%), the optimal RMSEs are obtained for ELM only on Concrete. The most optimal RMSEs is RESELM with 34 times, and RESELM ranks second in accuracy for most other cases. It can be seen that ELM, WELM and ORELM are most negatively affected as the outliers increase. In comparison, RESELM is hardly affected, which shows that RESELM can effectively suppress the effect of outliers.

This paper focuses on the improvement of RESELM in terms of robustness. In the case of lower outliers levels, the robustness of RESELM improves but does not achieve the optimal RMSE on all data sets, such as on the data sets Body-fat, which is mainly determined by the loss function, where the difference between the true value and the predicted value is larger, the bigger the corresponding loss function value. Researchers have proposed various loss functions in order to reduce the effect of outliers so that the robustness of the model will be better [18], [19]. As shown in Table I, the loss function of RESELM does not have a significant advantage in the lower outliers levels, so experiments with higher outliers levels are conducted to investigate the robustness of the proposed algorithm.

From Table I, it can be seen that the robustness of ELM remains uncompetitive in the case of higher outlier levels, and its RMSE increases with the outliers level. ELM applies a least squares loss function, and when the residuals are small, the least squares loss function does not differ significantly from the exponential squared loss function. Therefore, ELM can produce more accurate results when there are no outliers. However, this loss function increases infinitely with the growth of the residuals and is growing exponentially, so it cannot effectively constrain the effect of outliers, which makes the ELM less robust.

In the case of 25%-40% outliers levels, WELM does not obtain the optimal RMSE, but it has better accuracy than ELM. A weighted 2-norm loss function is proposed in WELM, which assigns different weights according to the value of the residuals to improve the robustness. However, the algorithm of WELM relies too much on the initial accuracy of the model, and the results obtained are not satisfactory.

ORELM has a better increase in accuracy than the lower outliers levels, and has the smallest difference from the optimal RMSE in most cases. ORELM uses the 1-norm loss function, which has a larger function value than the others when the residuals are small, and therefore has poorer accuracy. However, after the residuals increase, it has an advantage over the least squares loss function and the weighted 2-norm loss function. This is the reason that why ORELM is worse than IRWELM in most cases in the lower outliers levels but has better RMSEs than IRWELM in the cases of 25% to 40% outliers levels.



(a) Cooling

(b) Yacht

(c) Space

(d) NO2

(e) Heating

(f) Airfoil

(g) Abalone

(h) MG

Fig. 2. The performance of ELM, WELM, ORELM, IRWELM and RESELM on data sets with different outliers levels.

IRWELM is the opposite of ORELM. The accuracy of IRWELM ranks second 25 times in the lower outliers levels, and its ability to suppress outliers is worse than ORELM when the outliers level increases. IRWELM and WELM apply the same loss function, but IRWELM is more robust than WELM because IRWELM's solution method is an iterative reweighting algorithm. Compared with WELM which solves the output nodes directly, the iterative approach of IRWELM makes it obtain more accurate output nodes than those of WELM.

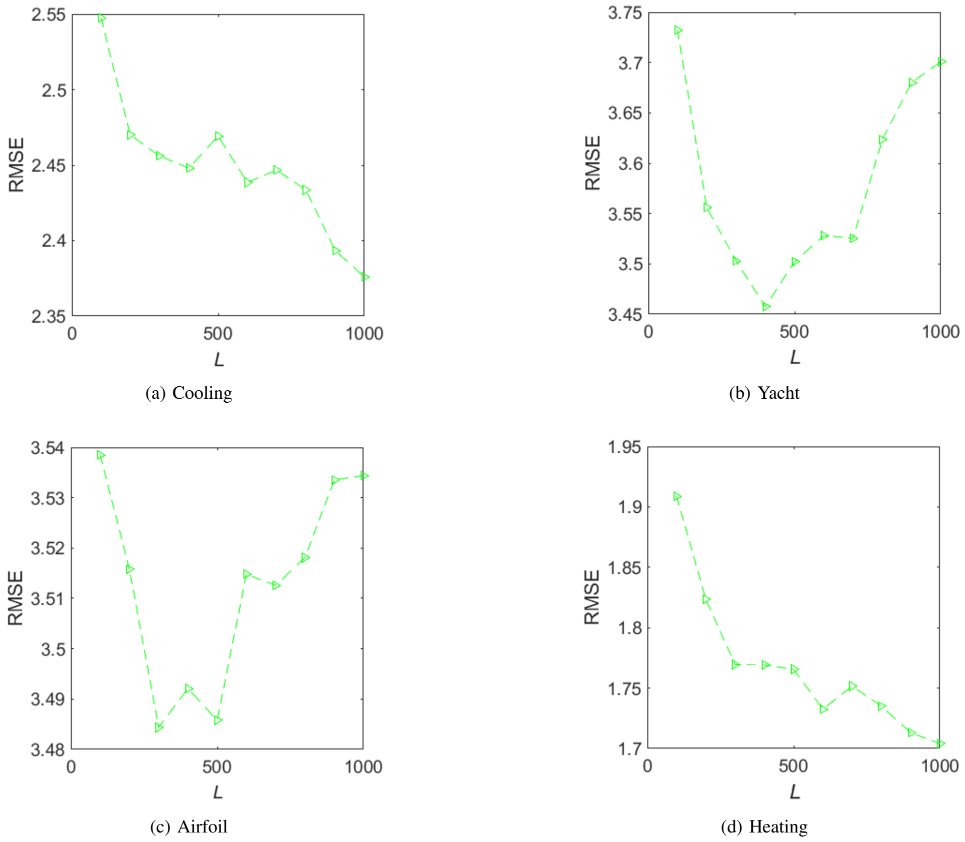
In the higher outliers levels, the RESELM has a significant advantage. The optimal RMSE is achieved 37 times due to the insensitive nature of the exponential squared loss function to higher outliers, it does not grow indefinitely, and the optimization problem is solved using an iterative approach in RESELM. Therefore, RESELM can effectively suppress outliers and has excellent generalizability.

Compared with ELM, WELM, ORELM, and IRWELM, RESELM obtains better accuracy than them in most cases. The RMSE is usually used in the study of improved models for ELM to illustrate the performance of the model in terms of accuracy, and experiments are conducted in [15], [16], [27] to verify the ability of the model to suppress outliers. The experiments in this paper show that the RMSE of RESELM is smaller than the other four models, so the proposed method effectively improves the robustness of ELM and has excellent generalization.

In Fig. 2, 8 benchmark data sets are chosen to show the variation of RMSE with the increasing outliers level for the five algorithms by line chart. It is more intuitive to observe the outliers level's effect on the algorithms' accuracy by the line chart. The line of ELM is almost always at the top of the axis and is most obvious on the Cooling, Yacht, Space, and Heating data sets, which have higher fold line from 0% outlier level to 40% outliers level than the other algorithms. WELM and IRWELM sometimes have worse RMSE than ELM in the higher outliers levels, as seen on the NO2, Airfoil, Abalone, and MG data sets. The fold line of ORELM are in the middle of ELM and RESELM on all data sets except the MG data set, which illustrates that the robustness of ORELM has improved somewhat compared to ELM but still suffer some effect in the higher outlier level compared to RESELM. The fold line of RESELM is always below all the folds except on the MG data set. Its accuracy is least affected by outliers, which means it has the best robustness.

B. Parameter Influence

Different parameters have an effect on the performance of the model. Next, the effect of hidden layer nodes L and the upper bound parameter σ of the exponential squared loss function on the performance of RESELM is examined. The experiments are conducted on four data sets without outliers, Cooling, Yacht, Airfoil, and Heating. In the experiments, the optimal parameters are selected for all parameters except the ones to be studied. The experiments reflect the effects of the parameters by RMSE.



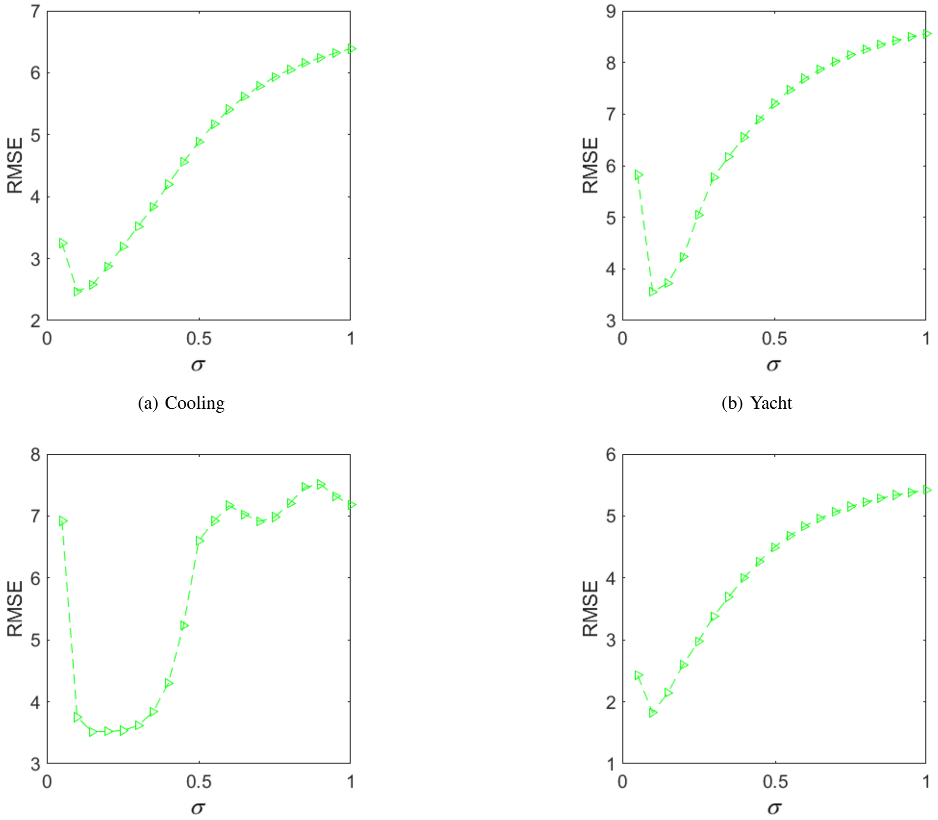
(a) Cooling

(b) Yacht

(c) Airfoil

(d) Heating

Fig. 3. The effect of L on RMSE.



(a) Cooling

(b) Yacht

(c) Airfoil

(d) Heating

Fig. 4. The effect of σ on RMSE.

There is no exact method to determine the number of hidden layer nodes L in ELM, and in practice previous experience or experimental methods within a certain range are usually used, but the number of L can affect the model's accuracy. When the L is too small, the model may have difficulty dealing with more complex problems. When L is too large, some nodes will not be meaningful to the model performance, which makes the model training time longer without improving the model accuracy. Therefore, the appropriate L is significant for the model's performance.

The effect of L on the model's accuracy is shown in Fig. 3, with L taking values from $\{100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$. The figure visualizes the effect of L. In Fig. 3(a) and (d), the curves of RMSE show a decreasing trend. The RMSE at L=1000 is the smallest, indicating that a minimum value is achieved in the given range. In Fig. 3(b) and (c), the RMSE decreases and then increases, and the optimal L is found in [300, 500]. The curves on four data sets are not smooth, which indicates that RESELM is a little sensitive to the network size. Therefore, when conducting experiments, choosing a more appropriate number of hidden layer nodes has an essential impact on the performance of RESELM.

To investigate the effect of σ on accuracy, experiments are conducted on four data sets using different σ chosen from the range [0, 1] with an interval of 0.05. In Fig. 4, the curves on the four data sets demonstrate that the RMSE first decreases and then continues to increase as σ increases. The global optimum is found in [0,0.5]. Although slightly different on the Airfoil data set, the global optimum is still in [0, 0.5]. The optimal RMSE can be obtained when the σ is small. By observing Fig. 1, it can be observed when the σ is small, the upper bound of the loss function also becomes smaller. When the residuals are larger, the value of the proposed loss function is also smaller for smaller σ, and the robustness of the model is better.

V. Conclusion

This paper proposes a robust ELM based on the exponential squared loss function (RESELM) for training samples contaminated with noise and outliers. The loss function used in the model is obtained based on correntropy. The nonconvexity of the exponential squared loss function enables RESELM to control the effect of outliers effectively. However, the nonconvexity also makes the model difficult to optimize. The proposed model is solved by formulating it as a DC programming and then adopting DCA. Experiments were conducted to verify the performance of RESELM on benchmark data sets with different outliers levels. The experimental results demonstrated that RESELM is non-sensitive to outliers and can obtain better robustness with a significant advantage in accuracy, especially in the case of 25% to 40% outliers levels. This paper discusses offline ELM, but in real-world problems, online learning is usually required, so future work considers extending RESELM to online sequential learning for better application to real-world problems.

Acknowledgment

The work was supported by the National Natural Science Foundation of China under Grant No. 61833005 and 61907033, the Postdoctoral Science Foundation of China under Grant No. 2018M642129.

References

[1] G. Huang, Q. Zhu, and C. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*, vol. 2, pp. 985–990, Ieee, 2004.

[2] G. Huang, Q. Zhu, and C. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[3] G. Huang, L. Chen, C. Siew, *et al.*, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.

[4] G. Huang, Q. Zhu, and C. Siew, "Real-time learning capability of neural networks," *IEEE Trans. Neural Networks*, vol. 17, no. 4, pp. 863–878, 2006.

[5] G. Huang, "Learning capability and storage capacity of two-hidden-layer feedforward networks," *IEEE transactions on neural networks*, vol. 14, no. 2, pp. 274–281, 2003.

[6] D. Zheng, Z. Hong, N. Wang, and P. Chen, "An improved lda-based elm classification for intrusion detection algorithm in iot application," *Sensors*, vol. 20, no. 6, pp. 1–19, 2020.

[7] J. Zeng, B. Roy, D. Kumar, A. S. Mohammed, D. J. Armaghani, J. Zhou, and E. T. Mohamad, "Proposing several hybrid pso-extreme learning machine techniques to predict tbm performance," *Engineering with Computers*, pp. 1–17, 2021.

[8] S. S. Chakravarthy and H. Rajaguru, "Automatic detection and classification of mammograms using improved extreme learning machine with deep learning," *Irbm*, vol. 43, no. 1, pp. 49–61, 2022.

[9] A. Boukerche, L. Zheng, and O. Alfandi, "Outlier detection: Methods, models, and classification," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–37, 2020.

[10] B. Frénay and M. Verleysen, "Classification in the presence of label noise: a survey," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 5, pp. 845–869, 2013.

[11] D. Nettleton, A. Orriols-Puig, and A. Fornells, "A study of the effect of different types of noise on the precision of supervised learning techniques," *Artificial intelligence review*, vol. 33, pp. 275–306, 2010.

[12] P. Meer, C. V. Stewart, and D. E. Tyler, "Robust computer vision: An interdisciplinary challenge," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 1–7, 2000.

[13] S. Mehrkanoon, X. Huang, and J. A. Suykens, "Non-parallel support vector classifiers with different loss functions," *Neurocomputing*, vol. 143, pp. 294–301, 2014.

[14] A. Ghosh, H. Kumar, and P. S. Sastry, "Robust loss functions under label noise for deep neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, 2017.

[15] W. Deng, Q. Zheng, and L. Chen, "Regularized extreme learning machine," *IEEE symposium on computational intelligence and data mining*, pp. 389–395, 2009.

[16] K. Zhang and M. Luo, "Outlier-robust extreme learning machine for regression problems," *Neurocomputing*, vol. 151, pp. 1519–1527, 2015.

[17] K. Chen, Q. Lv, Y. Lu, and Y. Dou, "Robust regularized extreme learning machine for regression using iteratively reweighted least squares," *Neurocomputing*, vol. 230, pp. 345–358, 2017.

[18] Y. Feng, Y. Yang, X. Huang, S. Mehrkanoon, and J. A. Suykens, "Robust support vector machines for classification with nonconvex and smooth losses," *Neural computation*, vol. 28, no. 6, pp. 1217–1247, 2016.

[19] X. Wang, K. Wang, Y. She, and J. Cao, "Zero-norm elm with non-convex quadratic loss function for sparse and robust regression," *Neural Processing Letters*, pp. 1–33, 2023.

[20] K. Wang, J. Cao, and H. Pei, "Robust extreme learning machine in the presence of outliers by iterative reweighted algorithm," *Applied Mathematics and Computation*, vol. 377, p. 125186, 2020.

[21] H. Pei, K. Wang, Q. Lin, and P. Zhong, "Robust semi-supervised extreme learning machine," *Knowledge-Based Systems*, vol. 159, pp. 203–220, 2018.

[22] W. Liu, P. Pokharel, and J. Principe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Transactions on signal processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[23] H. Xing and X. Wang, "Training extreme learning machine via regularized correntropy criterion," *Neural Computing and Applications*, vol. 23, pp. 1977–1986, 2013.

[24] L. An and P. Tao, "The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems," *Annals of operations research*, vol. 133, no. 1, pp. 23–46, 2005.

[25] R. Horst and N. Thoai, "Dc programming: overview," *Journal of Optimization Theory and Applications*, vol. 103, no. 1, pp. 1–43, 1999.

[26] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2011.

[27] P. Horata, S. Chiewchanwattana, and K. Sunat, "Robust extreme learning machine," *Neurocomputing*, vol. 102, pp. 31–44, 2013.

[28] T. Hodson, "Root mean square error (rmse) or mean absolute error (mae): when to use them or not," *Geoscientific Model Development Discussions*, vol. 15, no. 14, pp. 5481–5487, 2022.

Quality of Data (QoD) in Internet of Things (IOT): An Overview, State-of-the-Art, Taxonomy and Future Directions

Jameel Shehu Yalli[1], Mohd Hilmi Hasan[2], Nazleeni Samiha Haron[3], Mujeeb Ur Rehman Shaikh[4],
Nafeesa Yousuf Murad[5], Abdullahi Lawal Bako[6]
Computer and Information Sciences, Universiti Teknologi PETRONAS, Perak, Malaysia[1,2,3,4,5]
Department of Computing Science, University of Aberdeen, Scotland, United Kingdom (UK)[6]

Abstract—**The Internet of Things (IoT) data is the main component for finding the basis that allows decisions to be made intelligently and enables other services to be explored and used. Data originates from smart things that have the capabilities to connect and share data enormously with other things in the IoT ecosystem. However, the level of intelligence obtained and the type of services provided, all depend on whether the data is trusted or not. High-quality data is the most trusted;, it can be used to extract meaningful insights from an event and can also be used to provide good services to humans. Therefore, decisions based on high-quality and trusted data could be good, whereas those based on low-quality or untrusted data are not only bad but could also have severe consequences. The term Quality of Data (QoD) is used to represent data trustworthiness and is used throughout this paper. To the best of our knowledge, this work is the first to coin the term QoD. The problems that hinder QoD are identified and discussed. One if it is an outlier, it is a major feature of the data that degrades its overall quality. Several machine-learning techniques that detect outliers have been studied and presented, with few data-cleaning techniques. This paper aims to present the elements necessary to ensure QoD by presenting the overview of the IoT state-of-the-art. Then, data quality, data in IoT, and outliers are studied, and some quality assurance techniques that maintain data quality is presented. A comprehensive taxonomy is shown to provide state-of-the-art data in IoT. Open issues and future directions were suggested at the end of the paper.**

Keywords*—*Quality of Data (QoD); Internet of Things (IoT); RFID; WSN; Taxonomy; trustworthiness; outlier; anomaly; confusion matrix; QoD assurance technique

I. Introduction

The Internet of Things (IoT) has emerged as the new evolution of internet connecting different entities, things and objects from different sources around the globe, thereby generating enormous amounts of data every time, every second. The amount of data generated by the IoT is used and consumed by the objects with which it communicates than by humans. The number of servers needed to hold information for access by users is very large, giving an insight into the number of devices that connect to the internet. Then the number of IoT-connected things and devices is ten times the number of those internet PCs [**?**].

The tremendous amount of data generated by different things brings out the realization of big data problems, intelligent decision-making, and the development of many IoT applications. To start with the big data problem, when these things generate all the data and share, exchange, and store it in the cloud, the cloud centers need to provide enough storage to handle this data and also enough services to manage the data. Providing these two becomes a challenge for the cloud centers [**?**]. Advances in the technology have exploited many capabilities of these data from the things, thereby encouraging the continuous flow of data. Thus, IoT has become a major catalyst for big data problems.

For example, the scale of IoT is continuously expanding; reports show that by the end of the year 2025, the number of IoT connections could reach 24.6 billion, with a compound annual growth rate of 13% [**?**]. According to the International Data Corporation (IDC), there will be more than 38 billion linked things in 2025 and reach about 50 billion by 2030. Another projection by [**?**] reports that connections to IoT could be about 41 billion by 2025, which could generate approximately 79 zettabytes of data. According to [**?**], approximately 50 billion device connections exist today, with an estimated 75.44 billion device connections by 2025.

Intelligent decision-making is achieved when enough data is obtained from things covering enough scenarios and events to compare, deduce, and reach a conclusion. However, the IoT can perform all these based on the type and quality of data received; if the data is of good quality, decisions are likely to be good, but if the data lacks quality, decisions derived would also be bad [**?**]. Therefore, for a reliable, trusted, and intelligent decision, the data must be trustworthy.

Users and vendors found a lot of opportunities in the prevalence of IoT. New applications are being developed for the ease and comfort of the user. Some researchers are also working on AI applications that incorporate IoT, such as smart homes, smart cities, efficient energy management and distribution, and so on. To achieve optimality in IoT applications for both driven applications and network optimization, research has used meta-heuristic and heuristic algorithms to simulate physical and biological phenomena [**?**].

Examples of IoT applications in smart homes includes the adjustment of blinds according to temperature and environmental changes, the opening of doors for authorized vehicles, and the ordering of medical services when there is an emergency. In the traditional home, home devices are part of existing Internet expansion, but when IoT arrives, the migration of smart things begins to the IoT network [**?**]. When devices get corrupted, the consequences are severe. For example, when smart locks are hacked, anyone can access the home; when baby monitors are compromised. The homeowner can be scared; and when microwaves are hacked, it can cause fire. If the security of smart devices cannot to achieved, then smart homeowners may not want to live in their smart homes. On the contrary, they can expect to improve home safety by using intelligent surveillance services [**?**]. In addition, the privacy of smart homeowners must be preserved. However, the continuous incoming of data from smart devices can reveal the secrets of house owners. And this can pose serious threats to their privacy.

Some widely adopted applications include smart homes, smart cities, smart grids and smart transportation. IoT technologies have drastically changed our way of life [**?**]. The widespread interconnection of intelligent IoT objects distributed physically extends computational operation and communication costs to IoT objects with different specifications. These devices' sensor capabilities enable them to collect real-time data from the physical world. The analysis of such data enables us to build an intelligent world and make better decisions for its management. If these security concerns are not adequately addressed, the wide adoption of IoT applications will be severely hampered. Consider the two typical IoT application areas, Smart Home and Smart Health, where system-sensitive information and critical assets require high protection [**?**].

IoT will continue to affect our lives in many ways, both in our homes, offices, healthcare, cities, etc. IoT in our society can represent a symbolic capital of power [**?**]. The way to deal with this enormous amount of data has changed from manually entered data to autonomous devices such as RFID readers, sensor nodes, etc. Our common appliances could have embedded components to allow them to communicate and become more intelligent to ease our lives. Examples are the light bulb that warns you of its remaining life, a toaster that toasts bread and provides a weather forecast, a refrigerator, a television, a video camera, and a solar panel roof, which might all be IoT devices. Despite the comforts we enjoy when these appliances generate such data, it has posed a challenge to the servers to manage and process such a huge amount of data from around the globe, which leads to big data problems.

During the last decade, we have worried about computer protection. Last five years, we have been worried about our smartphones' protection, now we are worried about car protection; home appliances, wearables, and many other IoT devices. According to Hewlett-Packard, in 2014, 70% of the most common IoT devices were infected with serious vulnerabilities. The authors of [**?**] discussed various current security challenges: interoperability, resource constraints, the protection of privacy on the Internet of Things Security 297, and scalability. Thus, the security of IoT is currently the main concern and requires research attention.

IoT devices collect large amounts of data and transmit it to the network. There are many personal data in these data, such as blood pressure, pulse, electrocardiograms, place environment data, area humidity, room temperature, etc. Another authentication scenario is to consider the types of entities involved in the remote client and server scenarios [**?**]. Clients want to access servers' services. After the first registration, the client can have a mutual authentication with the server. After the authentication, the two can create a shared key, and the client can use this key to access the server's service. A server can provide its clients with different types of services. Servers have the responsibility to perform registration and password changes. Before these services can be provided to the client, the server must verify whether the client is registered or not [**?**].

We researched the quality of the data in this work and coined the term Quality (QoD), and to the best of our search and knowledge, we are the first to coin the term QoD. However, many factors contribute to the data inefficiency and lack of QoD. The first problems associated with the data and IoT devices include constraint capabilities, intermittent loss of connection, and deployment hazards [**?**]. Other problems come from smart things, such as node failure, faulty nodes, data loss, network congestion, architectural flaws, and so on [**?**]. A third-world problem is one created by humans and launched into the deployment field to gain some benefit; examples are side-channel attacks, node capture attacks, sensor impersonation, stolen verifier attacks, Sybil attacks, etc. For IoT to gain wide acceptance and embrace more deployment, the QoD needs to be ensured.

This survey first investigated the nature of the QoD, or data trustworthiness, in an IoT ecosystem. The data in IoT is explored further, from the data lifecycle to its characteristics and quality considerations in IoT, then the technology of RFID that allows the data to be shared is studied, and then down to the factors that affect the QoD in IoT. Some of the reliable techniques to ensure QoD are studied, and the techniques are presented in tabular form for ease of comparison. Data outliers, is the main component that compromises QoD, are researched, and the types and impacts of the outliers are presented for prevention and measurement. A comprehensive taxonomy that shows all the forms of data that can be used is designed for ease of understanding. Some IoT application domains, open issues, and future directions are presented as well.

The remainder of this article is presented as follows: after the introduction in Section 1, data in IoT is presented in Section II. Section III is the QoD assurance techniques and data outliers made in Section IV. A comprehensive data taxonomy is shown in Section V while some of the most common IoT application domains are made in Section VI. Open issues in QoD are presented in Section VII, and then, lastly, future directions and conclusion are in Section VIII. The paper has eight sections in all.

II. Overview of Data In IoT

Data is an important component that makes up the IoT paradigm and is the source of information and means of communication. Furthermore, QoD, or trustworthiness, is an essential requirement for any IoT ecosystem (i.e., IoT services). In the following sub-sections, we present the data life cycle in the context of the IoT. We also discuss the characteristics of IoT data. In addition, we discuss QoD in IoT. Moreover, we looked at RFID as the first technology on which the IoT is built, which allows data sharing among IoT devices. And then some of the factors that affect data quality in IoT were also discussed.

Since data is considered a valuable asset because of the insights gained about a phenomenon, it is used to provide intelligence in our daily lives dealings. Researchers, therefore, exploit the insights and intelligence in the data using different mining techniques and algorithms [?]. Data trustworthiness is essential in QoD to have a reliable handling of the data from the data itself, data interpretation, simulation results, and any other form of data representation. Data is characterized by losing its quality when some factors, such as things constrained resources, large-scale deployment, and intermittent connections are not obtained [?].

Some of these problems can be measured from the data quality dimensions that arise as a result of hazardous elements. One way to ensure that data quality is compromised is the identification of data outliers [?]. However, some outliers appear only to describe errors, while others describe rare events, e.g. unusually high temperature in a warehouse, maybe as a result of a fire. In IoT, QoD problems need to be solved. QoD is an essential requirement for any data user (things, entities, IoT services, and IoT user applications).

A. Data Lifecycle

In the original landscape setting of the internet, data primarily originates from users using their computers, surfing the web, engaging on social media networks, and generally being utilized to offer services to these users. In contrast, the IoT sees a paradigm shift where the majority of data is generated by interconnected devices, serving both as the source and primary recipient to deliver services to individuals. The Machine-to-Machine (M2M) is a precursor to IoT, which emphasizes data as the primary communication channel [?], facilitating autonomous collaboration among IoT objects to offer innovative services. Data holds significant value in the IoT, serving as a crucial asset that provides insights into various phenomena, individuals, or entities. These insights are leveraged by applications to deliver intelligent services ubiquitously. The accuracy of data is paramount, as any inaccuracies may compromise the reliability of extracted knowledge and subsequent actions based on it. Fig. 1 presents data life cycle stages.

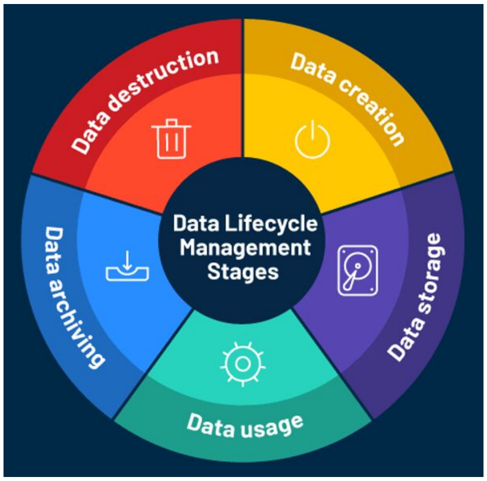


Fig. 1. Data life cycle stages.

B. Characteristics of IoT Data

The IoT device is embedded with a chip that can sense the environment and collect and share data with similar devices. The IoT devices are deployed mostly in hazardous environments, making them susceptible to natural effects such as earthquakes, rain, erosion, wind, etc. They can also be vulnerable to physical attacks and forced alteration by humans. These IoT sensors can be designed to measure variables of interest such as temperature, pressure, humidity, sleep habits, slope of a pipe, fitness level, movement positions, light intensity, and many more. However, some of the IoT characteristics are considered omnipresent, that is, erroneous, uncertain, noisy, distributed, voluminous, etc., while other characteristics can be considered dependent on the measured phenomenon, that is, continuous, smooth variation, periodicity, correlation, and Markovian behavior [?]. Some of these characteristics are:

1) Uncertain, noisy and erroneous data: uncertainty in QoD can make the data either incomplete, ignorant, ambiguous or imprecision caused by the constraint nature of the IoT nodes. Any factor that could make the data uncertain or put noise in the data, or make the data entirely wrong and have some wrong elements in it will compromise the QoD. And this can easily occur in any IoT ecosystem where the data is generated from volatile devices [?].

2) Voluminous and distributed data: In the IoT ecosystem, sensors can be deployed at any place to measure the parameters of interest. These sensors are densely deployed everywhere to gather enough data for decision-making and data management. The heterogeneous nature of IoT devices generates enormous amounts of data that are nearly impossible to manage and have challenges to manage. There is no standard IoT architecture to manage the total amount of data generated by its devices [?].

3) Smooth variation: in a continuous setting of data flow, such as the time stamp or time series, data is flown continuously at some interval. The collection and processing of such data requires some technique (like a machine learning technique) to collect the data and process it accordingly. An example is watering a tomato garden at some regular intervals [?].

4) Continuous data: The data here is similar to the smooth variation characteristic, but not necessarily that the data comes at a regular interval. The data can be random and have different patterns of arrival e.g., batch, stream, or real time. An example is to report any incident of traffic violation [?].

5) Correlation: The correlation feature exists in the IoT data set because of the heterogeneous nature of the IoT network. It consists of different sensors that measure different parameters. The data can have two correlations: spatial and temporal. When the data is correlated spatially according to the positions of the sensors, the processing of the data can give the best results in information form for certain phenomena. Whereas when the data is correlated temporally, the data might depend on its timestamp. For example, when the temperature values for the future are to be predicted, then the current temperature values can be used to make the prediction. It is also possible for the data set to have both types of correlations; the data can either be spatial (as related to memory space), temporal (as related to the time of its arrival), or both spatial and temporal [?].

6) Periodicity: This can be defined as the accuracy or age of specific data or the difference between the previous time stamp and the current time stamp as the item's punctuality or the data being sufficiently up to date for a task. Data sets that are related to scenarios may have inherently periodic patterns where the same values may occur at specific intervals [?].

7) Markovain behavior: An IoT sensor can be a function of a previous sensor, at a given time stamp of the previous sensor denoted by ti–1 [?].

C. Quality of Data in IoT

Quality of Data (QoD) in IoT can be seen as the possibility to ascertain the integrity of the data provided from its origin or the probability of the accuracy of the data [?]. Data must be clean, sensitized, and free from errors before it is transferred from a lower layer to an upper layer in the architectural stack. In other words, data must be reliable and trustworthy before it can be transmitted to preceding layers or peers for further processing. To compute trust in IoT ecosystem, it starts with the reliability of the sensors. Data from a reliable sensor could be considered trustworthy whereas data from a non-reliable sensor must be evaluated further to ascertain its correctness. However, the deployment nature of the IoT nodes in an unprotected environment makes the sensor vulnerable to attacks and unreliable [?].

Mechanisms have been developed to measure the trustworthiness in the WSN and the traditional internet; however, these mechanisms may not be suitable to ascertain the correctness of data in an IoT ecosystem due to the heterogeneous nature of the IoT, which is not the same as in WSN and the internet [?], [?]. Therefore, different mechanisms suitable for measuring QoD need to be developed, assessed, and implemented. Data is gathered massed from smart things providing ubiquitous services to users. For QoD to be ensured, the technologies used to allow the generation and sharing of the data should be addressed.

The first technologies used to allow things to connect and communicate were RFID, then WSN, which continues to evolve into other technologies up to the most widely used today, the internet (802.11 and its families). [?] assessed the quality of the data considering five dimensions, which include confidence, accuracy, timeliness, volume, and completeness. The new arising IoT applications rely on distributed and heterogeneous data for proper functioning; thereby, integration into IoT data is necessary [?]. However, maintaining such data becomes challenging due to the different sources it comes from [?], [?].

Data integrity is an essential asset in the authentication process of IoT devices. It ensures that node credentials are correct and unaltered. To ensure data integrity in an IoT distributed architecture, the MQTT protocol requires more attention because, when a connection is established between the nodes it has to transmit data to the destination, the credentials must be mutually verified to ensure that they have not been altered [?].

Assuming that the issuer ”P1” connects to the broker ”B1” directly, where the subscriber ”S1” connects to the broker ”B2” directly, the data is transferred from ”P1” to ”B1”, and also to ”B2” before it reaches the destination ”S1”. During this data transfer, the credentials of the data sender and the data recipient must be verified mutually. A common method of cryptography to ensure data integrity is the hash function (such as SHA-1 and SHA-2) [?].

However, ensuring QoD presumes fulfilling the criteria of accuracy, timeliness, precision, completeness, and reliability [?], [?], [?]. The authors of [?] define “timeliness as the data being current. That is, the most updated data in the most recent time”. While [?], sees timeliness from two different perspectives, i.e., “an error recovery of the data and its age item, this distinguishes the recorded timestamp from current system time while the regularity of the data is with respect to its application context”. Again, [?] defines “timeliness as the mean age value of the data in a source”. According to [?], timeliness is defined as “the extent to which data are sufficiently up-to-date for a task”.

D. Quality of Data in RFID

The first technology to be embraced by communicating entities is Radio Frequency Identification (RFID) [?]. The RFID system for authentication comprises three important tangible components: tags, readers, and data centers. The reader scans the tag to collect the necessary information and stores it in the data center. RFID can be seen as a transmitter microchip that is similar to an adhesive sticker. Active receives batteries that always emit data signals, while passive gets activated only when they are activated. The concept of radio technology was developed from RFID, where the chip does not have to view the reader in physical vision before it can communicate with it. While barcode technology requires the physical view of the reader to communicate with it [?]. RFID is also an actuator that stimulates events, an action a barcode could not do. WSN is a multi-strip wireless network connected to a dispersed sensor field that measures a specific data collection device's speed, humidity & temperature, whose values are transmitted to processing equipment. RFID is a short-range communication technology that is termed asymmetric, whereas WSN technology has a relatively long range and communication ability in a peer-to-peer fashion.

Since IoT's idea is to allow automatic connection and sharing of data among entities and any object with the ability to sense, process, transmit, and store information via the internet, it has made the network heterogeneous due to the different backgrounds of the objects. The early technologies start with RFID, then WSN, then Bluetooth, then wireless local area networks (WLANs), then wireless metropolitan area networks (WMANs), then cellular networks (LTE, 2G, 3G, 4G, 5G, and now 6G) [?], [?]. The IoT's vision is to [?] enable people and things or entities to communicate to anyone, at any time, anywhere through some sort of medium such as the internet [?]. With the rapid development of RFID technologies, Bluetooth, sensors, and smartphones, the applications and usage of IoT have increased tremendously, which directly affects daily life [?].

RFID automatically identifies entities, objects, and people. RFID's operation comes in three frequency ranges: the first is low frequency (LF), the second is high frequency (HF) and the third is ultra-high frequency (UHF). RFID devices can be separated into two groups: active and passive [?]. Active RFID requires an energy source, while passive RFID does not require any energy to power it. The encryption levels in RFID are: the first encryption mode with no traffic encryption, the second encryption mode with the Data Encryption Standard (DES), and the third encryption mode with AES-128 bits [?]. RFID uses radio waves to communicate with data in electronic devices to identify and sense the location around them.

In recent years, WSN and RFID have gained tremendous attention in the field of IoT applications. Both technologies have different capabilities and are used in different scenarios depending on their needs [?]. RFIDs provide reading content that is used to detect and identify objects they are associated with. WSNs provide dynamic content based on the environment in which they are installed [?]. However, these technologies are used as connectors between devices and local networks or even the wider Internet. This technology is necessary to identify devices, share information with each other, verify each other's identities, and broadcast other useful information to the network. In Table I, a list of some technologies used in the WSN and the internet are presented.

E. Factors Affecting Quality of Data

Data in the IoT is itself the weakest point due to many factors that affect its quality. When data lacks quality, it cannot represent the actual scenario it is being assigned to monitor, and it could have other negative effects both on the decision being made and the operational levels of any business or organization [?]. In order to have a phenomenon of interest, some potential problems in the IoT need to be addressed. Some of the issues facing the maintenance of QoD in the IoT include but are not limited to:

1) Resource constraints: Since their inception, IoT devices have been characterized as being resource constraints in nature. Because of the limited memory available in it, the small power consumption, and the lower computational ability of the IoT devices, it becomes difficult to trust all the data that comes from them especially if the data is more than what the device can hold, and naturally, more and more data keeps coming from these devices. The IoT devices are mostly battery-powered, and resources are scarce, data collection policies and trade-offs are inherently utilized to improve the quality and cleanliness of data [?].

2) Scalability: IoT is today being deployed on a global scale, starting from organizations to homes to cities and now to the globe. Any setting of the IoT deployment generates large amounts of data, and merging any setting or integrating any application makes the data even larger, thereby increasing the chances of error occurrence in the data [?].

3) Heterogeneity: IoT devices are heterogeneous in nature, they come from different settings and backgrounds the same way their data differs from the background it comes from. It is always more challenging to manage data of the same kind with data of different kind. IoT devices can only achieve functional optimality if they integrate heterogeneous data. Therefore, the issue of heterogeneity needs to be addressed perfectly [?].

4) Sensors: When sensors are deployed, they may suffer from a lack of accuracy in reporting their readings or from a loss of calibration. Some sensors may become faulty and then report incorrect data. This is a challenge that makes it mostly difficult to find the faulty sensor, especially in a large deployment setting [?].

5) Environment: The deployment is mostly in an unprotected surrounding affected by hazards and natural effects such as rain, earthquake, erosion, the mountain's summit, wind, or the intended attack by humans [?].

6) Network: The connection often gets lost and regained again due to limited resources, bad weather, infrastructural interference, and a bad signal. IoT is an IP network with a constrained higher loss of packets than the conventional network [?].

7) Vandalism: The environment is mostly unprotected and therefore suffers from physical attacks that include damage, stealing, altering, and forceful extraction of data from it. The vandalism also extends to animals whose aim is to search for food or scatter in any setting they come across. Therefore, this factor affects the QoD [?].

8) Dead node: It often happens in many circumstances that a node is dead, but data is continuously received from the node. This has made the quality of the data untrustworthy [?].

9) Privacy: This is a major part of the acceptance of IoT globally. People's data is not guaranteed to be secure, and when data is breached (like a patient's data), the damage is too high [?].

10)Data stream: Data from the smart IoT device is received and sent continuously in the back-end pervasive applications that use them [?].

Other problems include sleeplessness habits of some nodes, unauthorized access, altering the source code & attributes, incompleteness, etc. In the the memory devices could neither send large packets nor report events frequently due to constraints; therefore, only small-sized messages could be sent, which is insufficient to report all events. Also, the scarcity of resources will cause things to go into sleep mode to save energy. However, Internet Protocols (IP) maintain the backbone of IoT connectivity and are unsuitable for sleep modes so it requires the smart things to be operational at all times, unreliable readings, multi-source data inconsistencies & alignment, data duplication, data leakage, etc.

III. Data Outliers

A data outlier is an uncertainty in an event or scenario; it is a deviation from the normal distribution setting, resulting in problems in the data set and incorrect results in the model. Outliers are the major manifestations of QoD problems. Data outliers can also be defined as phenomena with extremely small chances of occurrence. It is again defined as points in a data set that is highly unlikely to happen in a given model [?].

In other words, anomalies are defined as patterns in data that do not conform to a well-defined notion of normal behavior [?]. Data outliers belong to a class of unreliable sets or groups; they fall outside of the normal status. In most machine learning models, they are considered the unusual points in a given data set [?]. Although these points have few chances

TABLE I. THE USE OF RFID & WSN TECHNOLOGY IN IOT, PROS AND CONS

N0	AUTHORS	DOMAIN	TECHNOLOGY	OBJECTIVE	PROS	CONS
1	[?]	WSN	RFID	To provide ultra-lightweight authentication by exploiting the RFID cache reader	It achieves reduced computational cost especially when authenticating a large number of tags, It achieves security	However, it needs to provide or expand storage space for a large number of tags
2	[?]	Big Data	RFID	To provide lightweight authentication based on simpler authentication protocols	The scheme combines multiple authentication protocols and runs well	Availability is guaranteed
3	[?]	WSN	WLAN	To develop a protocol needed to incorporate the TEPANOM solution and its architecture with the EAP infrastructure	This EAP supports many authentication mechanisms by introducing lower communication overhead compared to others, it does not require any global infrastructure, thus it is scalable	It cannot integrate closely the TEPANOM solution and its architecture with the EAP infrastructure
4	[?]	Smart Grid	6LowPAN, CoaP, IEEE, 802.15.4	To provide a lightweight authentication in the smart grid application	It maintains message integrity	The scheme is only tested and proved on a small scenario field nodes
5	[?]	WSN	GWN	To introduce a novel authentication and key agreement using bio hashing to eliminate false accept rate and false reject rate	Bio hashing has some functional advantages over biometrics such as high secure operation of imposter populations and genuine zero equal error rate level	The design is inefficient with limitations to support forward secrecy and unlinkability in two factor authentication. Also lack a dynamic identity mechanism to involve nonpublic key
6	[?]	WSN	GWN SN	To develop a lightweight biometric scheme to authenticate remote users and key agreement scheme for secure IoT services	User is authenticated remotely and offline	Memory requirements need to be found in the testbed and the lightweight feature extends to real IoT devices
7	[?]	IIoT	WSN	To design a lightweight computational biometric user authentication and key agreement scheme	The protocol is lightweight and less complex authentication is achieved. Authentication, availability, and integrity of data packets are guaranteed and the protocol needs further investigation	
8	[?]	WSN	WLAN	To design a light Weight node-to-node and node-to-node authentication protocols continuously	It authenticates each data transmitted between two nodes within a pre-defined time period in the IoT ecosystem	A more accurate model needs to be designed to minimize use of battery energy consumption and to discover more dynamic device features is challenging
9	[?]	NFC	RFID	To develop an Ultralight Weight Mutual Authentication Protocol to achieve forward security by using sub key and sub index number into its key update	Computational wise, the scheme is lightweight and proved to protect against synchronization attack	The NFC authentication needs to improve performance and function while still considering the security and privacy of the system
10	[?]	NFC	NFC	To develop a novel lightweight NFC identity authentication protocol for mobile NFC IoT networks	It has three working modes in the NFC mobile phone that it can work in the tag as the card, the reader, and support peer-to-peer file sharing	Its application electronic financial services still need privacy for public trust

to occur in the whole data set, but they often occur. Another formal metric-based outlier is the distance-based outlier (DB) [?]. It is defined when an object, let's say, 'x' in a given data set 'T', and the fraction of 'x' is higher than the distance 'D' within the context of 'DB', is considered an outlier. Fig. 2 shows an example of an outlier.

A. Types of Outliers

Data outliers as defined, are anomalies that do not conform with the normal behavior of the remaining data in the data set. Sometimes an outlier may represent an error, sometimes it may represent an incompatible element in a cluster, and sometimes it may even represent useful information. There are different types of data outliers in a model, but the most common ones are as follows:

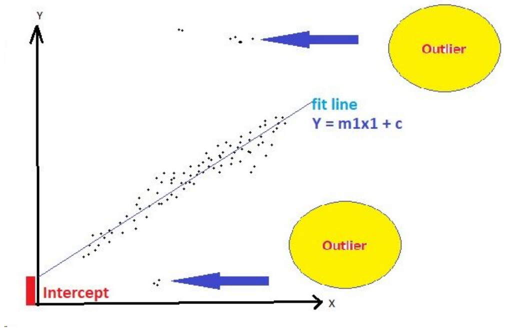


Fig. 2. Data outliers.

1) Error: This is any value generated as a result of node failure or node malfunction. When the node's battery is drained and is not replaced and the node continues to send data to the user's access point, it is very likely that the data is wrong or that repetitive data is sent. Sometimes the node may be altered by an attacker and forced to send the wrong data [**?**]. Attackers often try to extract some useful information from a captured node while trying to let it continue to send information to the base station. This process has already altered the normal process, so therefore, wrong data may be sent. It can also be affected by some natural effects, such as wind displacement, i.e., when a node is deployed in an environment and configured to measure a parameter of interest within a given ratio and the node is displaced outside that ratio, then the readings sent will not represent that area of interest [**?**].

2) Event: This is any scenario with an associated value generated due to a change in a certain setting or phenomenon. This can be demonstrated by a natural effect of event occurrence; for example, if in a hazardous environment, there is an accident or any natural phenomenon that changes the setting of the environment suddenly, then definitely the reading from the node at that particular time follows any sudden change as well, thereby not giving the expected outcome [**?**].

3) Point anomaly: This is the deviated data, a group of similar data that differ in value, behavior, and attributes. When usual data occurs in a given data set, deviating from the normal distribution pattern of the remaining data and the difference is huge enough to believe that it is out of context, then that outlier is considered a point anomaly [**?**]. For example; a model records the card withdrawals of an employee in Asia to occur once every day and then there appears to be a withdrawal transaction in Europe three hours from the last transaction in Asia, this translates into an impossible scenario; therefore, the data point of the transaction in Europe is classified as a clear point anomaly [**?**].

4) Contextual anomaly: This represents a value that could be an anomaly, but it does not depend on the context. Sometimes a deviation may occur, and still, the data may not be an outlier due to the context of the data set it belongs to [**?**]. Many scenarios happen where unusual data becomes an outlier in one context while being considered normal in another context. For example, given two data sets A and B, where data set A is a small data set in size, let's say with 100 rows, and data set B is large let's say it has 1000 rows. The calculation of its 'variance' using the same data point may become an outlier in data set A while proving normal in data set B [**?**].

5) Collective anomaly: represents a set of collected values that differs largely from other values in the data set. When more than one anomaly, let's say a group of anomalies appears in the data set, forming another cluster of anomalies, then it is referred to as a collective anomaly. This type of outlier is mostly identified in clustering algorithms such as the K-means algorithm, the Naive Bayes algorithm, the Decision Tree, etc. [**?**]. Fig. 3 gives examples of some of these outliers.

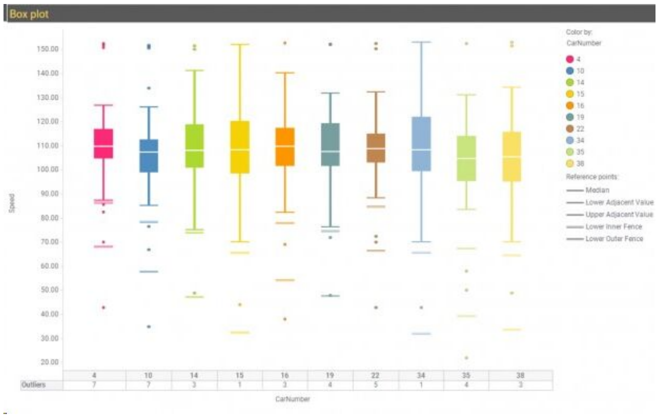


Fig. 3. Types of outliers.

B. An Outlier from the Confusion Matrix

A machine learning model is a powerful technique widely used to detect an outlier in a data set [**?**]. Today, there is seamless integration between the IoT domain and machine learning models. Many machine learning models are designed to solve IoT problems. One technique used to solve an IoT problem is the identification of an outlier by a model called the confusion matrix. This model has two true classes and two negative classes that are passed to the machine mode. An output is given by the model based on what it has learned. The four classes are described as follows:

1) True positive: when a true instance is passed to the model, the machine model computes the prediction and gives an output based on the data it learned. If the prediction corresponds to the actual or real event, then it is considered True Positive. For example, when it rains, the event is passed to the model, and the outcome is confirmed to be 'rain' or 'it rains'.

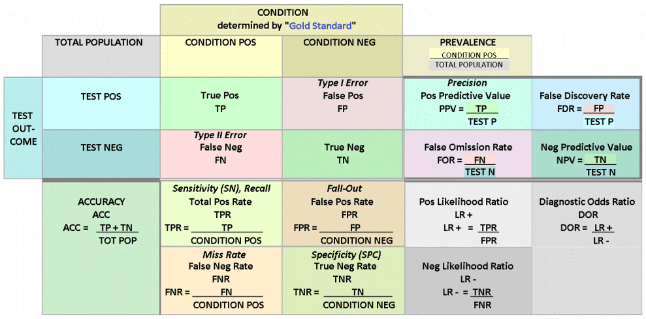
2) True negative: When the model receives the actual event and predicts wrongly giving an output that does not correspond to the expected outcome, then the scenario is considered True Negative. An example is when, in reality, it rains, the event is passed to the model, and the model predicts 'not to rain' or 'not raining'.

3) False positive: The other way around is to supply the model with the wrong event and expect it to learn and produce an output based on what it learns. When the outcome produced reveals the actual scenario that occurred while it was fed with the wrong event, then it is considered a False Positive. For example, when the status of an impregnated woman is passed to the model and it gives the output that the woman is pregnant then it is considered a False Positive.

4) False negative: But when the event passed to the model is the wrong event and the algorithm learns and predicts that the event is the wrong one, then it is called a False Negative. For example, if the model is fed with the woman's status as not being pregnant and produces the output as 'not pregnant'. Fig. 4 illustrates the confusion matrix in a diagram.

C. Impact of Outliers in the IOT

In IoT, environment data is obtained from the sensor as a result of measurement of parameters of interest, such as temperature, pressure, humidity, etc. and serves as an input to mine data so as to gain insights about a monitored phenomenon (e.g home, environment, health, etc. [**?**]. Based on these



		CONDITION determined by "Gold Standard"			
	TOTAL POPULATION	CONDITION POS	CONDITION NEG	PREVALENCE $\frac{\text{CONDITION POS}}{\text{TOTAL POPULATION}}$	
TEST OUT-COME	TEST POS	True Pos TP	*Type I Error* False Pos FP	*Precision* Pos Predictive Value PPV = $\frac{\text{TP}}{\text{TEST P}}$	False Discovery Rate FDR = $\frac{\text{FP}}{\text{TEST P}}$
	TEST NEG	*Type II Error* False Neg FN	True Neg TN	False Omission Rate FOR = $\frac{\text{FN}}{\text{TEST N}}$	Neg Predictive Value NPV = $\frac{\text{TN}}{\text{TEST N}}$
	ACCURACY ACC ACC = $\frac{\text{TP + TN}}{\text{TOT POP}}$	*Sensitivity (SN), Recall* Total Pos Rate TPR TPR = $\frac{\text{TP}}{\text{CONDITION POS}}$	*Fall-Out* False Pos Rate FPR FPR = $\frac{\text{FP}}{\text{CONDITION NEG}}$	Pos Likelihood Ratio LR + LR + = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic Odds Ratio DOR DOR = $\frac{\text{LR +}}{\text{LR -}}$
		Miss Rate False Neg Rate FNR FNR = $\frac{\text{FN}}{\text{CONDITION POS}}$	*Specificity (SPC)* True Neg Rate TNR TNR = $\frac{\text{TN}}{\text{CONDITION NEG}}$	Neg Likelihood Ratio LR - LR - = $\frac{\text{TNR}}{\text{FNR}}$	

Fig. 4. The confusion matrix.

insights, decisions can be made from different angles. It is clear that the conclusions reached from an erroneous data will produce bad and unsound decisions. For example, a model giving out too many false positives and false negatives such as a scenario of a campus fire alarm, the alarm system rings many times every week while in reality there is no cause for the panic [?].

Another scenario involves monitoring forest fires to respond quickly and take appropriate measures. In addition to component monitoring applications, the data on the state of the components should be reported accurately in order to protect expensive systems and avoid damage. An inability to provide accurate data can cause damage to whole systems or even the lives of people [?]. Other examples include forest fire alerting system that requires quick action, earthquake that requires quick evacuation to safe zones, etc.

The importance of accurate and reliable data is paramount, consider the examples above and imagine if any of the system do not alert about the occurrence of the dangers happening or the model reports otherwise about the danger, the consequences will be very severe and the lives of the people and entities is as high risk. When these nodes report actual data and the model predicts right and the system alarms correct, it will help and save lives and properties while if the nodes report faulty data, the model may easily predict wrong and the whole system may not function well to give the correct outcome, this is jeopardizing lives and properties.

The trustworthiness of the data is essential to the engagement of users and to the acceptance of IoT services and is therefore important to the success of the large-scale implementation of the IoT domain [?]. Data as a component of a holistic approach to managing IoT trust collection, reliability, and accuracy are the main concerns of data perception reliability.

As experiments and simulations are a good way to demonstrate and understand IoT systems likewise machine learning models are a good way to identify and prevent outliers in any given dataset. Many IoT experimentation test beds, such as the FIT-Equipex exist [?]. the authors of [?] examined several other existing testbeds (public and private). However, in order to study more on the impact of data anomalies. There are two real-world cases that examine the impact of QoD problems on the field of electronic health applications.

The first case study by [?] examines the effect of QoD problems on electronic health monitoring applications. This work identifies QoD issues that affect QoD criteria (e.g. accuracy, precision, timeliness, accessibility, and consistency) that are critical to providing appropriate help to the patient. There are three levels of data management defined to monitor cardiac scenarios: Data Acquisition, Data Processing, and Data Discovery. For example, at the level of data collection, the problems relate mainly to the performance of body sensors, the amount of data processed, and the quality of communications [?].

The second reported case study by [?] examines the poor impact of QoD on Ambient Assisted Living systems (AAL) systems, which results from the convergence of ambient intelligence and assisted living technologies. AAL supports monitoring applications (i.e. monitoring of health and well-being) to people in their homes. This paper argues that poor QoD alters the representation of events that occur, which hinders the system from giving appropriate support to users and causes incorrect reports on the health of patients, inefficient management of environmental conditions at home, etc. [?]. The application of e-health is one of the most important IoT applications taking into account the factors affecting human life and therefore tolerating uncertainty in the QoD.

IV. Quality of Data Assurance Techniques

In conventional programming, a common rule states: "Never trust user input," while in IoT the rule can be stated as: "Never trust things". This is proven as a result of uncertainties and inconsistencies in sensor data. In order to reduce the expensive effects of low QoD, a technique is needed to prepare data and improve its quality [?]. The following are five main techniques that could promise QoD in an IoT paradigm:

A. *Outlier detection*

This is to find the elements that differ from the normal distribution setting or deviate from the normal behavior of the data. The ultimate goal is to highlight outliers [?]. Identifying outlier detection in a model increases its overall reliability and efficiency. In addition, detecting outliers is the first step needed to handle all the events of inconsistencies. The next is the accuracy and reliability of data processing. If these are handled, then QoD will be ensured and better decisions will be made. Note that individual data element accuracy at the level itself does not increase because it relates only to the source of data generation and processing and cannot be improved [?]. The parameters used to detect outliers pay attention to highlighting data value differences to find out the outliers (for example, values that are not consistent with an established model) [?]. However, the QoD dimension values used to evaluate data are seen as insufficient, for example, the accuracy extracted from the measurement precision class of the sensor specified by the manufacturer. When the sensor fails to clean due to any cause, the accuracy becomes unreliable and irrelevant.

B. *Interpolation*

Interpolation is defined as a data generation method that can hence improve the QoD dimension of the data size (i.e., add the available data elements). However, interpolation is the opposite of completeness in effect, i.e., the ratio of the available data to non-interpolated items (i.e., both interpolated and non-interpolated) in the stream window in question. This scenario can be explained as an optimization effect to find the best compromise between these two dimensions, but nonetheless, is limited to satisfying user-defined QoD requirements [?]. Furthermore, when selecting interpolation techniques, it is important to consider the accuracy of the interpolation value [?], which is expected to meet user requirements. This technique involves the use of missing values based on a dataset. Data flows are described as missing data flow attributes or tuples (from sensor malfunctions and loss of connection) [?]. The missing data points represent gaps in the dataset that is available for a particular entity or a topic of interest. A model knowledge-generating processing of such a dataset, including those missing gaps in it would have incomplete knowledge and hence reduce QoD.

C. Data Integration

All the heterogeneous data from different landscapes must be integrated to overcome structural differences and inconsistencies and really benefit the universal service. Frameworks for Data Quality (DQ) techniques such as Resource Description Framework (RDF) and Web Ontology Language (OWL, 2009) provide standard mechanisms for data description to perform search, retrieval, and processing tasks more directly [?]. Also, linked data is a reliable approach to trigger data retrieval and integration in the IoT ecosystem. Another framework proposed by [?] integrates semantic data that uses the principles of Linked Data and semantic web technologies. The authors of [?] proposed an architectural model to integrate and incorporate all intelligent features into the smart application. Again, a Service Architecture Paradigm (SOA) model that extracts heterogeneity in intelligent objects and improves interoperability in the context of e-health applications is developed.

D. Data Deduplication

is a technique for the compression of data designed to lessen the volume of data stored by deleting duplicate data elements and replacing them with unique data references that remain unchanged. Data deduplication is a process of duplicating redundant data elements. It decreases the amount of data and affects the QoD of the data volume. [?] proposed a video duplication technique with considerations for privacy protection. The authors of [?] specify the deduplication technology for cloud-storage encrypted data. While [?] proposed a model for exploiting the deduplication capabilities together with the Hadoop framework.

E. Data Cleaning

The cleaning of data defines the life cycle of data; it starts with the selection of errors and goes down to the correction of identified errors and the identification of potential errors. It is also defined that the detection of anomalies is limited to the identification of anomalies, while data cleaning goes further to suppress the elements discovered. It has become a widely adopted technique for enterprise data management in data warehouses [?]. Data cleaning is a widespread topic of research in big data analysis [?]. It consists of three main stages: (i) determining the type of error; (ii) identifying potential errors; and (iii) correcting identified errors. It is also very common to manage enterprise data in the context of data storage. Fig. 5 presents some QoD techniques.

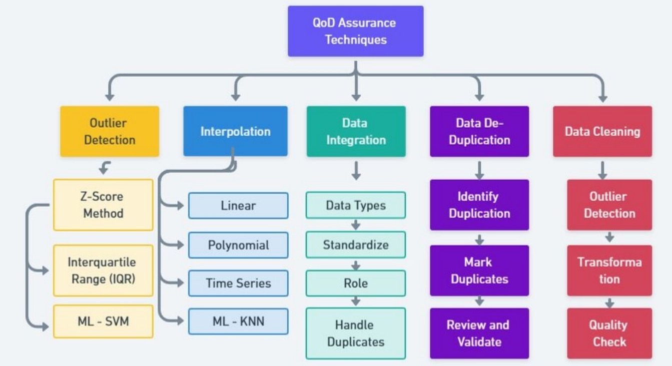


Fig. 5. QoD Assurance techniques.

V. Taxonomy for Quality of Data in IOT

The IoT and other similar domains such as WSN, have a set of attributes, features, characteristics, protocols, and technologies that are suited for their deployment and implementation. With a taxonomy, it can be seen what techniques have been used to measure data trustworthiness, what are the most important parameters to consider in ensuring QoD, what domains are integrated in the IoT data trustworthiness ecosystem, and what has been done and what needs to be done in the area.

A. Data Source

The data comes from one or more sources, and the source of the data determines whether the data can be trusted or not. Data could come from a sensor configured to provide the readings of an event, which is the source of the data, or from humans directly, especially in situations where interactive information is required by the model or the environment [?].

B. Data Processing

The first form is called stream or batch where data is usually sent at some specific time interval. The data in the same batch is mostly similar or has the same attributes. Data from the sensor could be sent in batch for processing, the data is first gathered and stored in the temporary memory of the device and then sent across the network to the server, or any base station provided. An example of batch processing can be any event whose action is not urgent [?]. Real-time processing is when the data is sensitive and needs to be processed immediately after it is obtained. In real-time, events such as fast decision-making making, banking transactions are examples of such scenarios. Another form of processing the data is computed in the near time, where the history of data is referred before making any computation. The processing of such data is not immediate, but time is also to be considered.

C. Data Type

The data obtained from the sensor could be presented to the user in the form of either Numeric, Alpha, Alpha Numeric, or even Symbols. The type of the data depends on the parameter measured and result representation [?].

D. Trust Type

The type of trust can be either direct or indirect. When data is gotten directly from a sensor it is considered as direct trust, likewise when the data is obtained from other sensors passing it to neighboring nodes, the quality of the data might degrade or the data may be intercepted and altered by natural phenomena, so therefore, the data is considered as indirect trust [?].

E. Trust Computation Location

The location of the node can be computed to determine the trustworthiness of the node. The computation can either be distributed or centralized. A cluster-based wireless sensor architecture can have the cluster head to compute the trustworthiness of the sensor node in the network, which is below a certain assigned value, where the actual data is sent to the gateway and then to the application layer. This considered as a centralized computation. Whereas in distributed trust computation, the node assesses the trustworthiness itself and send actual data to the cluster and gateway [?]

F. Trust Aggregation Method

is used to summarize trust evidence that has been gathered via nearby sensor node feedback or self-observations [?]. The majority of the literature's work use the following methods:

1) Weighted sum: This is the simplest model for aggregating trust scores. The method can summarise several factors that contribute to trust scores, the factors are multiplied by the specified weight, then adds up all results that contribute to a product that represents the trust score. Therefore, it is a commonly used technique for calculating trust score [?].

2) Bayesian inference: Due to its simplicity and solid statistical foundation, Bayesian inference is a popular confidence calculation model. This method considers trust as a random variable that follows a distribution of probability and which parameters are updated with new results.

3) Belief theory: The theory of belief, also known as the theory of evidence or Dempster-Shafer Theory (DST) provides a method for summarising confidence values from different pieces of evidence using Dempster's Rule of Combination. The rule assumes that this evidence is independent. Evidence is the confidence values computed by different sensor nodes in the network [?].

4) Regression analysis: Regression analysis is a method of aggregating confidence scores by calculating the relationships between data. The scores are calculated based on estimating the relationships of the trust factors and a number of other variables that affect the trust.

5) Fuzzy logic: This is a method that deals with estimation rather than fixed and exact conclusions, fuzzy logic also provides rules for reasoning. The confidence value determined with fuzzy logic can have a value between 0 and 1 with fuzzy measures [?].

6) Game theory: This involves making decisions between two or more decision-makers involved in certain conflicts or competitions. Game theory can be used to predict competitive rules of action with certainty. An example of using game theory models to ensure data reliability is the work of [19], which develops a defense strategy that ensures that sensor nodes are protected against attacks so that the difference between the value accepted by the sink and the true sense value is below a certain assigned value.

G. Trust Establishment

This refers to how to end a trust score from multiple properties. There are two aspects of the establishment of trust, namely single trust and multi-trust. A single trust implies that only one trust property is taken into account in the calculation of the total trust rating. On the contrary, multi-trust is a combination of trust and trust. Establishments use several trust factors to calculate the total data trust. Many proposed techniques utilize multi-trust factors to calculate trust scores, with two factors chosen on average. Among the factors used were communication, nodes' familiarity, energy, and nodes' reliability [?].

H. Trust Results

These are also called Trust Decisions, and are considered as an element of data trust calculation that deals with how the results are presented to the requestor or user. There are two options for representing the results either in binary or in a range of values or judgments. Binary represents the results as either trust or non-trust only. From this point of view, users or application layers can simply choose trusted data to process further. In terms of range, this means that the data reliability value calculated can fall within any range of possible degrees of trust. This is similar to the Likert scale, but the trust values can be determined by more than two options. As such, the requestor or user and application layers can decide accordingly on the basis of their decision logic [?].

I. Data QoS

The summation of all the packets involved in the transmission having the maximum delay (i.e., 400ms standard) and a regular jitter interval between its consecutive packets (i.e., 1ms interval) should then be able to produce the maximum number of packets received [?].

J. Node Quality

This is of two types: a resistant node which is able to prevent itself against side-channel attack and unclonability and any form of memory extraction. Most of such nodes have multiple ICs designed over one another to make it harder for an attack to recover anything from it. An unresistant node is one that is unable to protect itself from the attack mentioned [?].

K. Node State

A node can either be in a passive state where it remains ideal until it is triggered to send data, this node performs better in such a state since its battery will last longer while an active node is the one that is continuously measuring and sending data, the problem associated with this type of node is called sleep deprivation attack [?].

L. Measurable Parameters

These nodes are deployed to measure some readings called parameters. The parameters include but are not limited to the volume of an object, the temperature of a room or place, the pressure of equipment, humidity of a place, slope position of ground, fitness level of things, sleep habits in pipelines, state of the area, movement positions of liquids, etc. [?].

M. Data Accuracy

The data from the node is considered accurate as long as it is precise, correct, and reliable. The basis for having a sound decision is from accurate data while bad data produces a severe decision that comes with consequences. When data is accurate it is said to represent the actual scenario of an event [?].

N. Data Consistency

For the data to be consistent, it should satisfy standards and integrity, and the codes with which it is burned to the device must also be consistent in producing the actual result any time it is being run.

O. Data Completeness

Data is said to be complete if its record in the database is complete, the field data is complete and the single Unique Identifier is complete [?].

P. Data Timeliness

The timeliness or the regularity of the data is measured from its time stamp as well as its real-time updates [?].

Q. Data Relevance

Data is relevant according to the user requirements it satisfies and the contextual relevance of the data.

R. Data Robustness

When data is robust it should be able to accommodate fault tolerance by being resilient as well as quality monitoring [?].

S. Data Security

There are many forms to protect the data but for the sake of this research the most commonly practiced are access control mechanisms, encryption (symmetric and asymmetric), and data masking [?].

T. Metadata Quality

Metadata is the data from which data is made up or simply some supporting data that helps in the execution of the actual data. In Fig. 6, a comprehensive taxonomy for QoD in IoT is shown.

VI. IOT APPLICATION DOMAINS

- Smart Home: Smart homes have the vision to integrate intelligence into everyday objects such as appliances, door locks, surveillance cameras, garage doors, etc., and to communicate with existing cyber infrastructure [?]. Adding intelligence to physical objects is beneficial for improving people's lives, such as improving their comfort, convenience, security, and effective use of natural resources. For example, a smart home can adjust the blinds according to environmental changes, open garage doors automatically when an authorized vehicle approaches, or order medical services when there is an emergency. In smart homes, traditional home devices are part of existing Internet expansion. When devices are damaged, the consequences can be serious. For example, successfully hacking smart locks allows strangers to enter the house; compromising baby monitors can scare visitors away from the baby; hacking microwaves can cause a fire at home [?]. Smart homeowners may not want to live in smart homes if security is not guaranteed. On the contrary, they can expect to improve home safety by using intelligent surveillance services. In addition, the privacy of smart homeowners must be preserved. However, the continuous collection of data from smart home devices can reveal the private activities of house owners, as indicated in [?]. It poses serious threats to the privacy of owners.

- Another typical IoT application is the creation of smart networks, Smart grids are designed and implemented to improve the reliability, cost reduction, and efficiency of traditional power grid systems. It not only integrates green and renewable energy such as solar power, wind power, heat, etc. but also aims to improve the reliability and efficiency of traditional energy networks. Intelligent grid data communication networks connect many smart grid devices and play an essential role in achieving the above-mentioned objectives. It collects energy consumption data and monitors the state of smart grid systems. More applications can be developed based on smart and communication networks. For example, utilities can allocate and balance load more wisely based on energy use information collected. It can also help to design fair but scaled pricing models by taking into account unbalanced energy consumption in space and time. With a smart grid status monitoring application, you can identify faults in the grid system as quickly as possible and as well design new fault-tolerant mechanisms to better react to them. Many technologies, including Automatic Measurement Infrastructure (AMI), have been proposed to build smart network communication networks. Because so much data is moving around the mission-critical system, security is one of the most important concerns of such systems. Invading the smart grid [?] and cutting the supply of electricity to a large area can cause enormous physical and economic damage to society. Analyzing energy usage data can also reveal people's daily private activities.

- By embedding smart medical devices into the medical

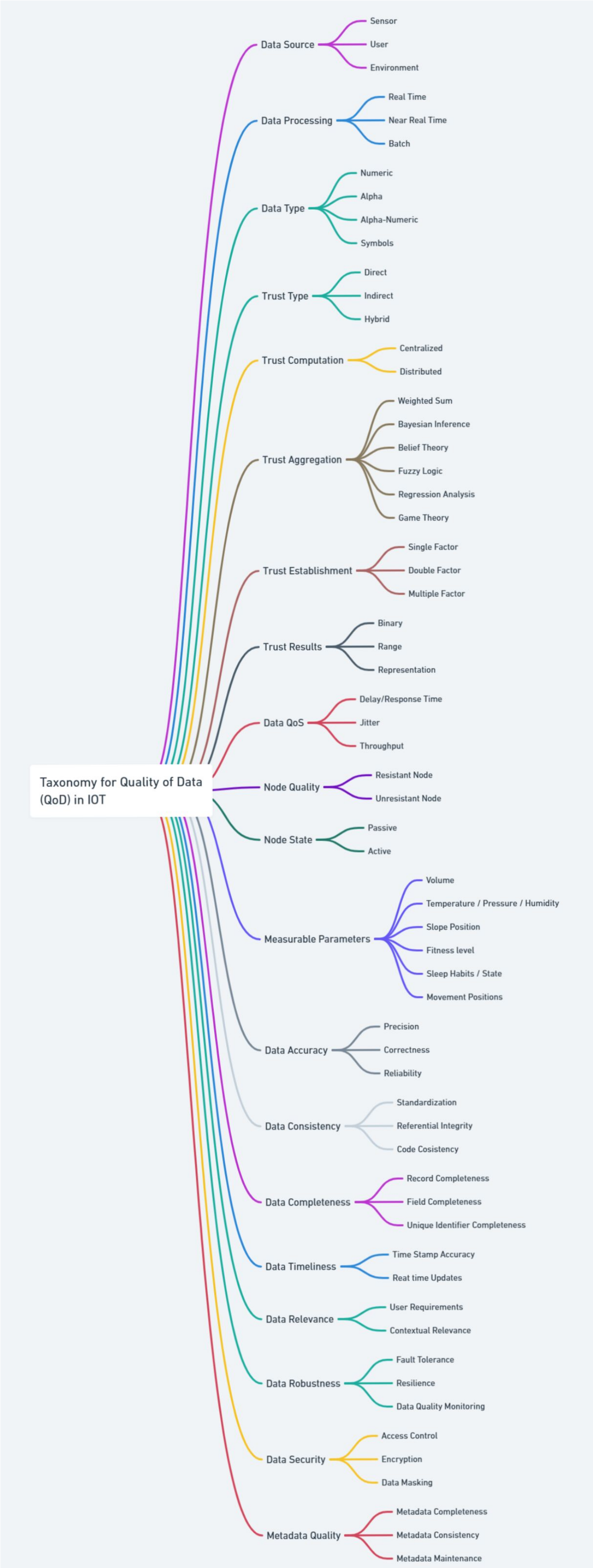


Fig. 6. Taxonomy of QoD in IOT.

infrastructure, health professionals can better monitor patients and use the data to determine those who need the most attention. In other words, healthcare professionals believe that prevention is more important and effective than treatment, so they can build proactive management systems based on collected data using the best of these networks of devices. Researchers also studied other possible techniques for implanting sensors into human bodies to monitor people's health status [?]. Using the collected data, health personnel can discover behavioral changes in the body of patients and medicines during treatment. Security is also an essential issue in Smart Connected Health. In network medical devices, data collection and monitoring of the status of the device is convenient, but there are also risks because instructions can be sent to terminate the device functions [?]. Stopping medical devices that are important to the patient's life such as heart injuries is extremely dangerous.

- Intelligent Transportation System (ITS) refers to the use of advanced technologies and data-based solutions to improve the efficiency, safety, and sustainability of transportation systems. Smart transportation uses technology and data to create more efficient, safer, and sustainable transportation systems that benefit individuals and communities. These systems play a key role in addressing urbanization, congestion, and environmental impacts in rapidly developing cities. In addition, IoT applications can manage passenger luggage at airports, and advise drivers about road conditions [?].
- The telecommunications industry includes a variety of technologies and services such as Global System for Mobile Communications (GSM) and Digital mobile network standards for voice and data communications on mobile phones and a few others. The Bluetooth technology is a short-range data exchange that is commonly used to connect headphones, speakers, and other peripheral devices. Wireless Local Area Network (WLAN): Wireless network technology allows devices to connect to the Internet or local network within a limited area. Wi-Fi called (wireless fidelity) is a service that allows voice calls and texts to be sent via Wi-Fi networks, providing coverage in areas with weak mobile signals [?]. Global Positioning System (GPS) is a satellite-based navigation system that provides accurate location information and is widely used in mobile phones, navigation devices, and vehicle tracking [?]. These technologies are essential to modern communication, connectivity, and location-based services.
- Logistics and Supply Chain Management: In logistics, RFID-embedded intelligent shelves enable real-time tracking of items, improving inventory visibility, accuracy, and efficiency [?]. This technology simplifies business, reduces errors, and helps companies optimize their supply chains to improve their performance and customer satisfaction. RFID-embedded smart shelves track items in real-time.
- Aerospace and Aviation Industry: In the aerospace and aviation industries, the Internet of Things (IoT) plays an essential role in enhancing safety, maintenance, and efficiency. Sensors and connected devices monitor aircraft components, collect performance data, allow real-time maintenance, improve reliability and safety, and reduce operational costs [?]. IoT has revolutionized the aerospace and aviation industries by providing real-time data and connectivity solutions to improve safety, reduce costs, improve passenger experience, and optimize the entire ecosystem, improve safety and security of elements.
- Automotive Industry: In the automotive industry, the IoT has changed the design, manufacture, operation, and maintenance of vehicles. It revolutionizes the automotive industry by making connected, autonomous, and safer vehicles, improving manufacturing processes, and improving consumer driving experiences. Sensors monitor and report vehicle parameters [?].

VII. Open Issues in Quality of Data for IOT

The following are a few issues that require additional research presented in four broad categories:

A. *Scalability*

The IoT can now be seen being deployed on a bigger scale, to say on an unprecedented scale that exceeds even the scale of the traditional Internet. Most of the solutions, however, are concentrated, and unlike distributed architectures, they do not provide sufficient flexibility and scaling for large-scale deployment [?].

B. *Heterogeneity of Data Sources*

The data generated in the IoT ecosystem comes from multiple types of objects, entities, sensors, RFID tags, etc. The architecture developed for IoT must be able to adapt to the heterogeneity of data origin. Furthermore, proposed technologies must be able to process different variables to meet the requirements of IoT applications. In order to meet the requirements of IoT applications, which may provide complex services based on multiple parameters such as user behavior, energy management, and home temperature relative to external temperature [?].

C. *Domain-agnostic/automated verification*

In IoT visions, things share data automatically with neighboring nodes based on their configuration. Domain-agnostic data cleaning methods confirm that data transmitted between "things" is uninterrupted, without human involvement, and with minimal human control, which is essential to the creation of a seamless IoT service [?].

D. *Distributed Architectures*

In addition to IoT scaling issues, distributed architectures also provide a platform that can adapt to faults and failure resilience. These functions are vital in the IoT perspective, because of the continuity and accessibility of data cleaning infrastructure, providing all-encompassing services even in the event of failures in ecosystems [?].

VIII. Conclusion and Future Directions

IoT offers great potential to connect millions of everyday objects and provide intelligent and ubiquitous services to help you live. The amount of data generated from the IoT infrastructure is very large. The collected data serves as the basis for obtaining insights that aid in decision-making, data management, and other services. QoD is an important interest in this scenario. Data security is linked to data quality, and the security of any model begins with data trustworthiness, which is essential to user participation and acceptance in the IoT paradigm. IoT is a promising domain, and there have been exciting results recorded in this field. In this context, QoD plays an important role. However, more research is needed to investigate how to improve QoD to ensure the widespread adoption and acceptance of IoT. Therefore, more work needs to be done to ensure effective and perfect decision-making, since data reliability is highly needed in IoT. The following are a few suggestions for future work to maintain QoD in IoT: Lightweight outlier detection Techniques, IoT network Traffic based Outlier Detection, Personalized QoD management platforms, QoD assessment-based outlier techniques, and QoD management middleware.

Acknowledgment

The authors would like to thank Yayasan Universiti Teknologi Petronas (YUTP) for providing the funding to carry out this research work. YUTP grant (015LC0-311) is granted by Universiti Teknologi PETRONAS (UTP).

References

[1] E. E. Broday and M. C. Gameiro da Silva, “The role of internet of things (iot) in the assessment and communication of indoor environmental quality (ieq) in buildings: a review,” *Smart and Sustainable Built Environment*, vol. 12, no. 3, pp. 584–606, 2023.

[2] G. R. C. de Aquino and C. M. de Farias, “Asclepius: Data quality framework for iot,” in *Proceedings of the Int’l ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications*, 2023, pp. 69–76.

[3] K. Fan, Y. Gong, C. Liang, H. Li, and Y. Yang, “Lightweight and ultralightweight rfid mutual authentication protocol with cache in the reader for iot in 5g,” *Security and Communication Networks*, vol. 9, no. 16, pp. 3095–3104, 2016.

[4] M. Saqib, B. Jasra, and A. H. Moon, “A lightweight three factor authentication framework for iot based critical applications,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 6925–6937, 2022.

[5] A. Kumar, R. Saha, M. Conti, G. Kumar, W. J. Buchanan, and T. H. Kim, “A comprehensive survey of authentication methods in internet-of-things and its conjunctions,” *Journal of Network and Computer Applications*, vol. 204, p. 103414, 2022.

[6] A. GhaffarianHoseini, N. D. Dahlan, U. Berardi, A. GhaffarianHoseini, and N. Makaremi, “The essence of future smart houses: From embedding ict to adapting to sustainability principles,” *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 593–607, 2013.

[7] I. Rouf, H. Mustafa, M. Xu, W. Xu, R. Miller, and M. Gruteser, “Neighborhood watch: Security and privacy analysis of automatic meter reading systems,” in *Proceedings of the 2012 ACM conference on Computer and communications security*, 2012, pp. 462–473.

[8] X. Pan, Z. Ling, A. Pingley, W. Yu, N. Zhang, and X. Fu, “How privacy leaks from bluetooth mouse?” in *Proceedings of the 2012 ACM conference on Computer and communications security*, 2012, pp. 1013–1015.

[9] K. Sha, W. Wei, T. A. Yang, Z. Wang, and W. Shi, “On security challenges and open issues in internet of things,” *Future generation computer systems*, vol. 83, pp. 326–337, 2018.

[10] Z. Liu, K.-K. R. Choo, and M. Zhao, “Practical-oriented protocols for privacy-preserving outsourced big data analysis: Challenges and future research directions,” *Computers & Security*, vol. 69, pp. 97–113, 2017.

[11] L. Nataliia and F. Elena, “Internet of things as a symbolic resource of power,” *Procedia-Social and Behavioral Sciences*, vol. 166, pp. 521–525, 2015.

[12] R. Chetan and R. Shahabadkar, “A comprehensive survey on exiting solution approaches towards security and privacy requirements of iot,” *International Journal of Electrical and Computer Engineering*, vol. 8, no. 4, p. 2319, 2018.

[13] M. Nair, S. Dang, and M. A. Beach, “Iot device authentication using self-organizing feature map data sets,” *IEEE Communications Magazine*, 2023.

[14] E. E.-D. Hemdan, Y. M. Essa, M. Shouman, A. El-Sayed, and A. N. Moustafa, “An efficient iot based smart water quality monitoring system,” *Multimedia Tools and Applications*, pp. 1–25, 2023.

[15] T. C. C. Nepomuceno, V. D. H. de Carvalho, K. T. C. Nepomuceno, and A. P. C. Costa, “Exploring knowledge benchmarking using time-series directional distance functions and bibliometrics,” *Expert Systems*, vol. 40, no. 1, p. e12967, 2023.

[16] S. M. Tahsien, H. Karimipour, and P. Spachos, “Machine learning based solutions for security of internet of things (iot): A survey,” *Journal of Network and Computer Applications*, vol. 161, p. 102630, 2020.

[17] R. Huo, S. Zeng, Z. Wang, J. Shang, W. Chen, T. Huang, S. Wang, F. R. Yu, and Y. Liu, “A comprehensive survey on blockchain in industrial internet of things: Motivations, research progresses, and future challenges,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 88–122, 2022.

[18] N. ALBAZZAI, O. RANA, and C. PERERA, “Camera as a sensor towards augmenting anomaly detection in internet of things systems: A survey.”

[19] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, “Outlier detection for temporal data: A survey,” *IEEE Transactions on Knowledge and data Engineering*, vol. 26, no. 9, pp. 2250–2267, 2013.

[20] N. Haron, J. Jaafar, I. A. Aziz, M. H. Hassan, and M. I. Shapiai, “Data trustworthiness in internet of things: A taxonomy and future directions,” in *2017 IEEE conference on big data and analytics (ICBDA)*. IEEE, 2017, pp. 25–30.

[21] S. Sagar, A. Mahmood, K. Wang, Q. Z. Sheng, J. K. Pabani, and W. E. Zhang, “Trust–siot: Towards trustworthy object classification in the social internet of things,” *IEEE Transactions on Network and Service Management*, 2023.

[22] H. Foidl and M. Felderer, “An approach for assessing industrial iot data sources to determine their data trustworthiness,” *Internet of Things*, vol. 22, p. 100735, 2023.

[23] S. Sagar, A. Mahmood, Q. Z. Sheng, J. K. Pabani, and W. E. Zhang, “Understanding the trustworthiness management in the social internet of things: a survey,” *arXiv preprint arXiv:2202.03624*, 2022.

[24] M. Alabadi, A. Habbal, and X. Wei, “Industrial internet of things: Requirements, architecture, challenges, and future research directions,” *IEEE Access*, 2022.

[25] L. Wei, Y. Yang, J. Wu, C. Long, and B. Li, “Trust management for internet of things: A comprehensive study,” *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7664–7679, 2022.

[26] F. M. R. Junior and C. A. Kamienski, “A survey on trustworthiness for the internet of things,” *IEEE Access*, vol. 9, pp. 42 493–42 514, 2021.

[27] Z. N. Aghdam, A. M. Rahmani, and M. Hosseinzadeh, “The role of the internet of things in healthcare: Future trends and challenges,” *Computer methods and programs in biomedicine*, vol. 199, p. 105903, 2021.

[28] L.-A. Tang, X. Yu, S. Kim, Q. Gu, J. Han, A. Leung, and T. La Porta, “Trustworthiness analysis of sensor data in cyber-physical systems,” *Journal of Computer and System Sciences*, vol. 79, no. 3, pp. 383–401, 2013.

[29] G. Han, J. Jiang, L. Shu, J. Niu, and H.-C. Chao, “Management and applications of trust in wireless sensor networks: A survey,” *Journal of Computer and System Sciences*, vol. 80, no. 3, pp. 602–617, 2014.

[30] D. Hui-hui, G. Ya-jun, Y. Zhong-qiang, and C. Hao, "A wireless sensor networks based on multi-angle trust of node," in *2009 International Forum on Information Technology and Applications*, vol. 1. IEEE, 2009, pp. 28–31.

[31] A. Klein and W. Lehner, "Representing data quality in sensor data streaming environments," *Journal of Data and Information Quality (JDIQ)*, vol. 1, no. 2, pp. 1–28, 2009.

[32] ——, "How to optimize the quality of sensor data streams," in *2009 Fourth International Multi-Conference on Computing in the Global Information Technology*. IEEE, 2009, pp. 13–19.

[33] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the internet of things: perspectives and challenges," *Wireless Networks*, vol. 20, pp. 2481–2501, 2014.

[34] S.-M. Luo, Z.-J. Ge, Z.-W. Wang, Z.-Z. Jiang, Z.-B. Wang, Y.-C. Ouyang, Y. Hou, H. Schatten, and Q.-Y. Sun, "Unique insights into maternal mitochondrial inheritance in mice," *Proceedings of the National Academy of Sciences*, vol. 110, no. 32, pp. 13 038–13 043, 2013.

[35] K. Gupta and K. Yadav, "Data collection method to improve energy efficiency in wireless sensor network," in *International Conference of Advance Research and Innovation (ICARI-2015)*, 2015.

[36] S. Shitole and A. Gujar, "Securing broker-less publisher/subscriber systems using cryptographic technique," in *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*. IEEE, 2016, pp. 1–6.

[37] J. E. Bailey and S. W. Pearson, "Development of a tool for measuring and analyzing computer user satisfaction," *Management science*, vol. 29, no. 5, pp. 530–545, 1983.

[38] C. Batini, A. Rula, M. Scannapieco, and G. Viscusi, "From data quality to big data quality," *Journal of Database Management (JDM)*, vol. 26, no. 1, pp. 60–82, 2015.

[39] W. S. Geisler, "Contributions of ideal observer theory to vision research," *Vision research*, vol. 51, no. 7, pp. 771–781, 2011.

[40] T. Dasu and T. Johnson, *Exploratory data mining and data cleaning*. John Wiley & Sons, 2003.

[41] C.-C. Lai, T.-C. Wang, C.-M. Liu, and L.-C. Wang, "Probabilistic top-^{k} dominating query monitoring over multiple uncertain iot data streams in edge computing environments," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8563–8576, 2019.

[42] X. Jia, Q. Feng, T. Fan, and Q. Lei, "Rfid technology and its applications in internet of things (iot)," in *2012 2nd international conference on consumer electronics, communications and networks (CECNet)*. IEEE, 2012, pp. 1282–1285.

[43] A. Khan, A. Ahmad, M. Ahmed, J. Sessa, and M. Anisetti, "Authorization schemes for internet of things: requirements, weaknesses, future challenges and trends," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 3919–3941, 2022.

[44] H. Chen, X. Jia, and H. Li, "A brief introduction to iot gateway," in *IET international conference on communication technology and application (ICCTA 2011)*. IET, 2011, pp. 610–613.

[45] M. Kumhar and J. Bhatia, "Emerging communication technologies for 5g-enabled internet of things applications," *Blockchain for 5G-Enabled IoT: The new wave for Industrial Automation*, pp. 133–158, 2021.

[46] S. S. Sabry, N. A. Qarabash, and H. S. Obaid, "The road to the internet of things: a survey," in *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*. IEEE, 2019, pp. 290–296.

[47] S. Ijaz, M. A. Shah, A. Khan, and M. Ahmed, "Smart cities: A survey on security concerns," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 2, 2016.

[48] A. W. Nagpurkar and S. K. Jaiswal, "An overview of wsn and rfid network integration," in *2015 2nd International Conference on electronics and communication systems (ICECS)*. IEEE, 2015, pp. 497–502.

[49] K. Pal, "Challenges of using wireless sensor network-based rfid technology for industrial iot applications," *Handbook of Research on Advancements of Contactless Technology and Service Innovation in Library and Information Science*, pp. 80–100, 2023.

[50] G. Mudra, H. Cui, and M. N. Johnstone, "Survey: An overview of lightweight rfid authentication protocols suitable for the maritime internet of things," *Electronics*, vol. 12, no. 13, p. 2990, 2023.

[51] G. B. Mohammad, S. Shitharth, S. A. Syed, R. Dugyala, K. S. Rao, F. Alenezi, S. A. Althubiti, and K. Polat, "Mechanism of internet of things (iot) integrated with radio frequency identification (rfid) technology for healthcare system," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–8, 2022.

[52] A. Kumar, K. Gopal, and A. Aggarwal, "Cost and lightweight modeling analysis of rfid authentication protocols in resource constraint internet of things," *Journal of Communications Software and Systems*, vol. 10, no. 3, pp. 179–187, 2014.

[53] M. P. Pawlowski, A. J. Jara, M. J. Ogorzalek *et al.*, "Compact extensible authentication protocol for the internet of things: enabling scalable and efficient security commissioning," *Mobile Information Systems*, vol. 2015, 2015.

[54] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. S. Shen, "A lightweight message authentication scheme for smart grid communications," *IEEE Transactions on Smart grid*, vol. 2, no. 4, pp. 675–685, 2011.

[55] J. Srinivas, S. Mukhopadhyay, and D. Mishra, "Secure and efficient user authentication scheme for multi-gateway wireless sensor networks," *Ad Hoc Networks*, vol. 54, pp. 147–169, 2017.

[56] P. K. Dhillon and S. Kalra, "A lightweight biometrics based remote user authentication scheme for iot services," *Journal of Information Security and Applications*, vol. 34, pp. 255–270, 2017.

[57] B. Khalid, K. N. Qureshi, K. Z. Ghafoor, and G. Jeon, "An improved biometric based user authentication and key agreement scheme for intelligent sensor based wireless communication," *Microprocessors and Microsystems*, vol. 96, p. 104722, 2023.

[58] Y.-H. Chuang, N.-W. Lo, C.-Y. Yang, and S.-W. Tang, "A lightweight continuous authentication protocol for the internet of things," *Sensors*, vol. 18, no. 4, p. 1104, 2018.

[59] K. Fan, P. Song, Y. Yang *et al.*, "Ulmap: Ultralightweight nfc mutual authentication protocol with pseudonyms in the tag for iot in 5g," *Mobile Information Systems*, vol. 2017, 2017.

[60] K. Fan, C. Zhang, K. Yang, H. Li, and Y. Yang, "Lightweight nfc protocol for privacy protection in mobile iot," *Applied Sciences*, vol. 8, no. 12, p. 2506, 2018.

[61] C. Liu, P. Nitschke, S. P. Williams, and D. Zowghi, "Data quality and the internet of things," *Computing*, vol. 102, no. 2, pp. 573–599, 2020.

[62] R. Perez-Castillo, A. G. Carretero, M. Rodriguez, I. Caballero, M. Piattini, A. Mate, S. Kim, and D. Lee, "Data quality best practices in iot environments," in *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. IEEE, 2018, pp. 272–275.

[63] L. Zhang, D. Jeong, and S. Lee, "Data quality management in the internet of things," *Sensors*, vol. 21, no. 17, p. 5834, 2021.

[64] R. Kollolu, "A review on wide variety and heterogeneity of iot platforms," *The International journal of analytical and experimental modal analysis, analysis*, vol. 12, pp. 3753–3760, 2020.

[65] D. Sehrawat and N. S. Gill, "Smart sensors: Analysis of different types of iot sensors," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2019, pp. 523–528.

[66] S. L. Ullo and G. R. Sinha, "Advances in smart environment monitoring systems using iot and sensors," *Sensors*, vol. 20, no. 11, p. 3113, 2020.

[67] K. Gulati, R. S. K. Boddu, D. Kapila, S. L. Bangare, N. Chandnani, and G. Saravanan, "A review paper on wireless sensor network techniques in internet of things (iot)," *Materials Today: Proceedings*, vol. 51, pp. 161–165, 2022.

[68] X. Yang, L. Shu, Y. Liu, G. P. Hancke, M. A. Ferrag, and K. Huang, "Physical security and safety of iot equipment: A survey of recent advances and opportunities," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4319–4330, 2022.

[69] M. Shafiq, H. Ashraf, A. Ullah, M. Masud, M. Azeem, N. Jhanjhi, and M. Humayun, "Robust cluster-based routing protocol for iot-assisted smart devices in wsn." *Computers, Materials & Continua*, vol. 67, no. 3, 2021.

[70] P. M. Chanal and M. S. Kakkasageri, "Security and privacy in iot: a survey," *Wireless Personal Communications*, vol. 115, no. 2, pp. 1667–1693, 2020.

[71] B. M. Alencar, R. A. Rios, C. Santana, and C. Prazeres, "Fotstream: A fog platform for data stream analytics in iot," *Computer Communications*, vol. 164, pp. 77–87, 2020.

[72] M. A. Samara, I. Bennis, A. Abouaissa, and P. Lorenz, "A survey of outlier detection techniques in iot: review and classification," *Journal of Sensor and Actuator Networks*, vol. 11, no. 1, p. 4, 2022.

[73] A. Gaddam, T. Wilkin, and M. Angelova, "Anomaly detection models for detecting sensor faults and outliers in the iot-a survey," in *2019 13th International Conference on Sensing Technology (ICST)*. IEEE, 2019, pp. 1–6.

[74] M. A. Bhatti, R. Riaz, S. S. Rizvi, S. Shokat, F. Riaz, and S. J. Kwon, "Outlier detection in indoor localization and internet of things (iot) using machine learning," *Journal of Communications and Networks*, vol. 22, no. 3, pp. 236–243, 2020.

[75] H. Ghallab, H. Fahmy, and M. Nasr, "Detection outliers on internet of things using big data technology," *Egyptian Informatics Journal*, vol. 21, no. 3, pp. 131–138, 2020.

[76] M. V. Brahmam and S. Gopikrishnan, "Nodstac: Novel outlier detection technique based on spatial, temporal and attribute correlations on iot bigdata," *The Computer Journal*, p. bxad034, 2023.

[77] P. D. Rosero-Montalvo, Z. István, P. Tözün, and W. Hernandez, "Hybrid anomaly detection model on trusted iot devices," *IEEE Internet of Things Journal*, 2023.

[78] C. Karras, A. Karras, and S. Sioutas, "Pattern recognition and event detection on iot data-streams," *arXiv preprint arXiv:2203.01114*, 2022.

[79] A. Gaddam, T. Wilkin, M. Angelova, and J. Gaddam, "Detecting sensor faults, anomalies and outliers in the internet of things: A survey on the challenges and solutions," *Electronics*, vol. 9, no. 3, p. 511, 2020.

[80] E.-S. Apostol, C.-O. Truică, F. Pop, and C. Esposito, "Change point enhanced anomaly detection for iot time series data," *Water*, vol. 13, no. 12, p. 1633, 2021.

[81] A. Shahraki and Ø. Haugen, "An outlier detection method to improve gathered datasets for network behavior analysis in iot," 2019.

[82] J. Liang, "Confusion matrix: Machine learning," *POGIL Activity Clearinghouse*, vol. 3, no. 4, 2022.

[83] D. Gupta *et al.*, "Prediction of sensor faults and outliers in iot devices," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2021, pp. 1–5.

[84] L. Boukela, G. Zhang, M. Yacoub, S. Bouzefrane, S. B. B. Ahmadi, and H. Jelodar, "A modified lof-based approach for outlier characterization in iot," *Annals of Telecommunications*, vol. 76, pp. 145–153, 2021.

[85] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for internet of things," *Journal of network and computer applications*, vol. 42, pp. 120–134, 2014.

[86] G. Z. Papadopoulos, A. Gallais, G. Schreiner, and T. Noel, "Importance of repeatable setups for reproducible experimental results in iot," in *Proceedings of the 13th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, 2016, pp. 51–59.

[87] C. O'Reilly, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Anomaly detection in wireless sensor networks in a non-stationary environment," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1413–1432, 2014.

[88] S. Rodríguez-Valenzuela, J. A. Holgado-Terriza, J. M. Gutiérrez-Guerrero, and J. L. Muros-Cobos, "Distributed service-based approach for sensor data fusion in iot environments," *Sensors*, vol. 14, no. 10, pp. 19 200–19 228, 2014.

[89] J. McNaull, J. Augusto, M. Mulvenna, and P. McCullagh, "Ambient assisted living systems and technologies: a data and information quality perspective," *ACM Transactions on Data and Information Quality*, vol. 4, no. 1, pp. 1–15, 2012.

[90] J. McNaull, J. C. Augusto, M. Mulvenna, and P. McCullagh, "Data and information quality issues in ambient assisted living systems," *Journal of Data and Information Quality (JDIQ)*, vol. 4, no. 1, pp. 1–15, 2012.

[91] ——, "Multi-agent system feedback and support for ambient assisted living," in *2012 Eighth International Conference on Intelligent Environments*. IEEE, 2012, pp. 319–322.

[92] R. Togneri, G. Camponogara, J.-P. Soininen, and C. Kamienski, "Foundations of data quality assurance for iot-based smart applications," in *2019 IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE, 2019, pp. 1–6.

[93] M. K. Abiodun, J. B. Awotunde, R. O. Ogundokun, E. A. Adeniyi, and M. O. Arowolo, "Security and information assurance for iot-based big data," in *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*. Springer, 2021, pp. 189–211.

[94] M. Bures, T. Cerny, and B. S. Ahmed, "Internet of things: Current challenges in the quality assurance and testing methods," in *International conference on information science and applications*. Springer, 2018, pp. 625–634.

[95] C. A. Ardagna, E. Damiani, J. Schütte, and P. Stephanow, "A case for iot security assurance," *Internet of Everything: Algorithms, Methodologies, Technologies and Perspectives*, pp. 175–192, 2018.

[96] J. Buelvas, D. Múnera, D. P. Tobón V, J. Aguirre, and N. Gaviria, "Data quality in iot-based air quality monitoring systems: a systematic mapping study," *Water, Air, & Soil Pollution*, vol. 234, no. 4, p. 248, 2023.

[97] D. Li, L. Yan, Y. Liu, Q. Yin, S. Guo, and H. Zheng, "Data quality improvement method based on data correlation for power internet of things," in *2019 12th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2. IEEE, 2019, pp. 259–263.

[98] H. Khodkari, S. Maghrebi, and R. Branch, "Necessity of the integration internet of things and cloud services with quality of service assurance approach," *Bulletin de la Société Royale des Sciences de Liège*, vol. 85, no. 1, pp. 434–445, 2016.

[99] A. M. Nagib and H. S. Hamza, "Sighted: a framework for semantic integration of heterogeneous sensor data on the internet of things," *Procedia Computer Science*, vol. 83, pp. 529–536, 2016.

[100] R. Morabito, R. Petrolo, V. Loscrí, and N. Mitton, "Enabling a lightweight edge gateway-as-a-service for the internet of things," in *2016 7th International Conference on the Network of the Future (NOF)*. IEEE, 2016, pp. 1–5.

[101] S. Li, L. D. Xu, and S. Zhao, "The internet of things: a survey," *Information systems frontiers*, vol. 17, pp. 243–259, 2015.

[102] X. Huang, P. Craig, H. Lin, and Z. Yan, "Seciot: a security framework for the internet of things," *Security and communication networks*, vol. 9, no. 16, pp. 3083–3094, 2016.

[103] R. Sethi, B. Bhushan, N. Sharma, R. Kumar, and I. Kaushik, "Applicability of industrial iot in diversified sectors: evolution, applications and challenges," *Multimedia technologies in the Internet of Things environment*, pp. 45–67, 2021.

[104] X. Ding, H. Wang, G. Li, H. Li, Y. Li, and Y. Liu, "Iot data cleaning techniques: A survey," *Intelligent and Converged Networks*, vol. 3, no. 4, pp. 325–339, 2022.

[105] T. Wang, H. Ke, X. Zheng, K. Wang, A. K. Sangaiah, and A. Liu, "Big data cleaning based on mobile edge computing in industrial sensor-cloud," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1321–1329, 2019.

[106] J. Guo and R. Chen, "A classification of trust computation models for service-oriented internet of things systems," in *2015 IEEE International Conference on Services Computing*. IEEE, 2015, pp. 324–331.

[107] J. S. Yalli and M. H. Hasan, "A unique puf authentication protocol based fuzzy logic categorization for internet of things (iot) devices," in *Proceedings of the 2023 12th International Conference on Software and Computer Applications*, 2023, pp. 246–252.

[108] H.-S. Lim, G. Ghinita, E. Bertino, and M. Kantarcioglu, "A game-theoretic approach for high-assurance of data trustworthiness in sensor networks," in *2012 IEEE 28th International Conference on Data Engineering*. IEEE, 2012, pp. 1192–1203.

[109] N. Mohamed, *Critical Socio-Technical issues surrounding mobile computing*. IGI Global, 2015.

[110] J. S. Yalli, S. B. Abd Latif, A. H. A. Hashim, and M. K. Alam, "An improved qos in the architecture, model and huge traffic of multi-media applications under high speed wireless campus network," 2006.

[111] J. Yalli, S. Latif, M. Masud, M. Alam, and A. Abdallah, "A comprehensive analysis of improving qos and imm traffic of high speed wireless campus network," in *2014 IEEE Symposium on Computer*

Applications and Industrial Electronics (ISCAIE). IEEE, 2014, pp. 12–17.

[112] J. S. Yalli, S. B. Abd Latif, and S. Bari, "Interactive multi-media applications: Quality of service guaranteed under huge traffic," *International Journal of Computer Applications*, vol. 105, no. 7, 2014.

[113] F. A. Garba, K. I. Kunya, Z. A. Zakari *et al.*, "A proposed novel low cost genetic-fuzzy blockchain-enabled internet of things (iot) forensics framework," *Scientific and practical cyber security journal*, 2021.

[114] M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, "Data-stream-based intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study," *IEEE Systems journal*, vol. 9, no. 1, pp. 31–44, 2014.

[115] G. Leroy, H. Chen, and T. C. Rindflesch, "Smart and connected health [guest editors' introduction]," *IEEE Intelligent Systems*, vol. 29, no. 3, pp. 2–5, 2014.

[116] M. Rahman, B. Carbunar, and M. Banik, "Fit and vulnerable: Attacks and defenses for a health monitoring device," *arXiv preprint arXiv:1304.5672*, 2013.

[117] S. Muthuramalingam, A. Bharathi, S. Rakesh Kumar, N. Gayathri, R. Sathiyaraj, and B. Balamurugan, "Iot based intelligent transportation system (iot-its) for global perspective: A case study," *Internet of Things and Big Data Analytics for Smart Generation*, pp. 279–300, 2019.

[118] A. M. Al-Momani, M. A. Mahmoud, and M. S. Ahmad, "Factors that influence the acceptance of internet of things services by customers of telecommunication companies in jordan," *Journal of Organizational and End User Computing (JOEUC)*, vol. 30, no. 4, pp. 51–63, 2018.

[119] A. M. Luthfi, N. Karna, and R. Mayasari, "Google maps api implementation on iot platform for tracking an object using gps," in *2019 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*. IEEE, 2019, pp. 126–131.

[120] Y. Song, F. R. Yu, L. Zhou, X. Yang, and Z. He, "Applications of the internet of things (iot) in smart logistics: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4250–4274, 2020.

[121] G. Karakuş, E. Karşıgil, and L. Polat, "The role of iot on production of services: A research on aviation industry," in *Proceedings of the International Symposium for Production Research 2018 18*. Springer, 2019, pp. 503–511.

[122] M. A. Rahim, M. A. Rahman, M. M. Rahman, A. T. Asyhari, M. Z. A. Bhuiyan, and D. Ramasamy, "Evolution of iot-enabled connectivity and applications in automotive industry: A review," *Vehicular Communications*, vol. 27, p. 100285, 2021.

[123] C. Campolo, G. Genovese, G. Singh, and A. Molinaro, "Scalable and interoperable edge-based federated learning in iot contexts," *Computer Networks*, vol. 223, p. 109576, 2023.

[124] I. Bedhief, M. Kassar, and T. Aguili, "Empowering sdn-docker based architecture for internet of things heterogeneity," *Journal of Network and Systems Management*, vol. 31, no. 1, p. 14, 2023.

[125] J. Gaskin, A. Elmaghbub, B. Hamdaoui, and W.-K. Wong, "Deep learning model portability for domain-agnostic device fingerprinting," *IEEE Access*, 2023.

[126] F. Azzedin and T. Alhazmi, "Secure data distribution architecture in iot using mqtt," *Applied Sciences*, vol. 13, no. 4, p. 2515, 2023.